Normalized unitary synaptic signaling of the hippocampus and entorhinal cortex predicted by deep learning of experimental recordings

Keivan Moradi¹,4, Zainab Aldarraji², Megha Luthra², Grey P. Madison³ & Giorgio A. Ascoli¹,²✉

Biologically realistic computer simulations of neuronal circuits require systematic data-driven modeling of neuron type-specific synaptic activity. However, limited experimental yield, heterogeneous recordings conditions, and ambiguous neuronal identification have so far prevented the consistent characterization of synaptic signals for all connections of any neural system. We introduce a strategy to overcome these challenges and report a comprehensive synaptic quantification among all known neuron types of the hippocampal-entorhinal network. First, we reconstructed >2600 synaptic traces from ∼1200 publications into a unified computational representation of synaptic dynamics. We then trained a deep learning architecture with the resulting parameters, each annotated with detailed metadata such as recording method, solutions, and temperature. The model learned to predict the synaptic properties of all 3,120 circuit connections in arbitrary conditions with accuracy approaching the intrinsic experimental variability. Analysis of data normalized and completed with the deep learning model revealed that synaptic signals are controlled by few latent variables associated with specific molecular markers and interrelating conductance, decay time constant, and short-term plasticity. We freely release the tools and full dataset of unitary synaptic values in 32 covariate settings. Normalized synaptic data can be used in brain simulations, and to predict and test experimental hypothesis.
The discovery of place cells and grid cells underscored the importance of the hippocampal formation as a key neural substrate for spatial navigation\textsuperscript{1,2}, fueling an intensive investigation of this brain region. Understanding spatial coding requires a model of the information flow in the underlying cellular circuit. Synapses mediate neuronal communication by enabling the transmission of signal from the axon of a (sender) neuron to the dendrite or perisomatic area of a (receiver) neuron. An electrical signal is thus recordable from the postsynaptic cell upon activation of the presynaptic cell. Different synapses produce distinct signals ultimately orchestrating behavior and cognition\textsuperscript{3}. For example, the release probability, conductance, and short-term plasticity (ST-P) vary among pairs of neuron types\textsuperscript{4}. Plastic changes in synaptic signaling subserve adaptive processes underlying memory. Identifying aberrant synaptic dynamics is crucial to the elucidation of the pathophysiology of diseases such as schizophrenia and depression\textsuperscript{5,6}. Yet, the synaptic physiology of most neuronal connections remains poorly understood.

The summed synaptic activity of multiple contacts connecting two neurons is a unitary signal. Unitary synaptic signals are typically measured by paired recording\textsuperscript{7}, allowing for post-hoc identification of both presynaptic and postsynaptic neuronal types. Unfortunately, paired recordings are based on blind probing with a low success rate in finding connected pairs. Accordingly, sample sizes for this method are typically small. Collating recordings from different studies may increase statistical power if they can be mapped to a common framework. Such a framework needs to unify synaptic measurement methods (synaptometrics), experimental conditions such as temperature and slicing preparation methods, and classification of neuronal types.

The knowledge base Hippocampome.org provides a useful starting point by identifying 122 neuron types based on their main neurotransmitter (glutamate or GABA), their dendritic and axonal morphologies, and their molecular expression. Since neuronal connections require the anatomical co-location of a presynaptic axon and a postsynaptic dendrite (or soma), synapses could be classified based on the morphological patterns of the corresponding neurons\textsuperscript{8}. Specifically, if a neuron type sends its axons to an anatomical subregion and layer in which another neuron type extends its dendrites, these two neurons can make a connection. The set of all axonal-dendritic co-locations is then trimmed by excluding experimentally refuted connectivity to yield a list of potential connections. Thus, anatomical constraints and known connection specificities are used to reduce the number of potential connections from all 14,884 (122 × 122) pairs of neuron types to only 3,120 (∼21%) in the rodent hippocampus and entorhinal cortex\textsuperscript{9,10}. Nevertheless, Hippocampome.org lacked until now a quantitative description of normalized synaptic physiology for all potential connections in the circuit.

To coalesce synaptic physiology data from the hippocampal formation, we mined approximately 1,200 publications, annotated more than 2,600 synaptic electrophysiology traces or values, extracted the synaptometrics, annotated the recording methods and experimental conditions, and mapped the data to the neuron types and potential connections of Hippocampome.org\textsuperscript{11}. However, the data are in various formats, requiring unification into a common formalism. This can be achieved using a phenomenological description of synaptic dynamics that summarizes synaptic properties in a low dimensional parametrization of the ground truth\textsuperscript{12–14}. In such an approach, synaptic amplitude is defined by a conductance (g), decay kinetics with a deactivation time constant (τ_d), and short-term plasticity (ST-P) through the dynamics of synaptic resource utilization and recovery determined by three parameters: a recovery time constant (τ_r), a facilitation time constant (τ_f), and the utilization ratio (U).

Large g values lead to high synaptic amplitudes, and large τ_d values result in slow kinetics. Depending on the calcium concentration in the presynaptic terminal, each synaptic event increases resource utilization rate, reflecting the number of released neurotransmitters and postsynaptic receptors occupied at any moment. U determines the utilization increment proportion after each event, but it is not the only factor. Resource utilization rate diminishes between events as calcium is reabsorbed. The utilization reduction pace is determined by τ_r. When τ_r is large, utilization reduction speed is slow, and synapses have a higher probability to facilitate. Since synaptic resources are limited, utilization may cause depletion. Therefore, synapses could have fewer resources for the next event, unless they recover quickly. The factor τ_f determines the recovery speed. High τ_f indicates lower recovery rate which makes synapses more likely to undergo short-term depression.

Quantifying synaptic physiology with the aforementioned parameters enables the unification of diverse experimental data. Nevertheless, different covariates, including species, sex, temperature, and recording modality, make it challenging to compare synapses beyond the scope of the original studies (Fig. 1). Published reports also do not cover all potential connections. Synaptic data in Hippocampome.org are only available for ∼84% of potential connections in the hippocampal formation. Moreover, due to the often-ambiguous identification of cell types, each synaptic signal is typically mappable to several potential connections\textsuperscript{11}. To solve these problems, the mined data require proper integration\textsuperscript{15–17}. Specifically, a comprehensive data model is needed to normalize existing data and infer missing information. Deep learning is a powerful tool for data integration and supports multi-target regression\textsuperscript{18–21}. In fact, trial-to-trial heterogeneity may increase the robustness of machine learning\textsuperscript{22}. Despite its successes in other fields, deep learning has never been employed to integrate synaptic electrophysiology data.

This study introduces a strategy to normalize unitary synaptic properties and employs it to generalize the available electrophysiology data by inferring the missing information. First, we effectively reconciled data collected through multifarious methods by fitting the quantitative measurements of recorded connections with a parametric synapse model. Then, we trained a predictive deep learning model to normalize the data for covariates and validated the prediction accuracy against the measured experimental variability. The model can infer missing values in arbitrary conditions and resolve ambiguous neuronal identities. Thus, for the first time, we comprehensively analyzed the normalized synaptic properties of all potential connections of the rodent entorhinal–hippocampal network and unraveled crucial factors governing synaptic physiology.

Results

We compiled, digitized, and reconstructed from the published literature a comprehensive dataset of 2,621 synaptic signals recorded from the dentate gyrus, CA3, CA2, CA1, subiculum, and entorhinal cortex\textsuperscript{11}. For each recording, we annotated the detailed experimental conditions with 75 covariates (Methods; Table 1) and mapped the potential pair of presynaptic and postsynaptic neuron types among 3,120 potential connections identified by Hippocampome.org\textsuperscript{9,10}. While this synaptic database constitutes a uniquely information-rich resource, its quantitative analysis requires solving distinct challenges (Fig. 2). First, researchers record synaptic signals in different modalities (current- or voltage-clamp) and widely diverse experimental conditions, which cannot be directly compared. Second, synaptic measurements can rarely be ascribed to single identified presynaptic and postsynaptic neuron types: in most cases, the mapping is ‘fuzzy’ and
matches several potential connections (green arrows in Fig. 2). Third, synaptic data are unavailable for a sizeable minority of potential connections. Additionally, certain experiments only include one synaptic event (e.g., the upper right signal in Fig. 2), thus providing no information on short-term plasticity. We fit all synaptic recordings to the same model via signal simulation to solve part of the first challenge (normalizing recording modality and a subset of covariates). To solve the remaining challenges (normalizing the rest of the covariates, disambiguating potential connections, and inferring missing data), we devise an original strategy based on machine learning.

Modeling comparable synaptic parameters from diverse measures and modalities. Data integration starts with the digitization of published synaptic recordings (Fig. 3a). These signals are diverse in terms of measurement modalities (current vs voltage) and the composition of intracellular and extracellular solutions affecting reversal potentials ($E_{rev}$). To transform these data into a comparable form, we fitted all digitized signals to a simplified Tsodyks, Pawelzik, and Markram (TPM) model, representing synaptic properties with five parameters (Methods)\textsuperscript{12,13}. These synapse-specific parameters ($g$, $\tau_d$, $\tau_r$, $\tau_f$, and $U$) depend on the combination of presynaptic and postsynaptic neuronal types involved and are estimated by fitting the TPM model output to the digitized signals (Fig. 3b). The model also requires a small set of measurements that depend on experimental settings and the properties of the postsynaptic neuron: $E_{rev}$, the initial value of the membrane voltage ($V_m$), membrane time constant ($\tau_m$), and capacitance ($C_m$). We corrected the signals before parametric fitting to eliminate the impact of processes causing slow signal fluctuations (Suppl. Figure 1a-b and Methods). The fitting quality was satisfactory resulting in minimal optimization error (Suppl. Figure 1c-left). However, we could find a weak correlation between paired-pulse ratio and the optimization error suggesting the TPM model simulates depressing synapses better than facilitating synapses (Suppl. Figure 1c-right). The TPM model produced comparable synaptic parameters and normalized the data with respect to synaptic driving force ($V_m - E_{rev}$) by converting synaptic amplitudes to conductance. Overall, the process reduces data dimensionality by describing every signal with only five values.

Construction and validation of a predictive model of all synapses. The fitted parameters for matching potential connections in different experimental conditions reveal a large degree of variation that could be associated with covariates such as animal...
Table 1 List of features used for machine learning.

| Features                                      | Dimensions | Columns |
|-----------------------------------------------|------------|---------|
| Potential presynaptic neuron(s)               | 122 neuron types (one-hot encoded) | 122     |
| Potential postsynaptic neuron(s)              | 122 neuron types (one-hot encoded) | 122     |
| Stimulation method                            | Evoked, unitary, spontaneous, or miniature | 3       |
| Type of response                              | GABAergic or glutamatergic | 1       |
| Response contamination with another neurotransmitter | True vs false | 1       |
| GABA$_A$ or NMDA response or block            | Intracellular block, extracellular block, pure slow response, contaminated with AMPA or GABA$_A$, and discoverable but was absent | 5       |
| Preservation of intracellular monoamines      | True vs false | 1       |
| Calcium-permeable AMPA density                | High, low, absent, or blocked | 4       |
| $E_{rev}$                                      | AMPA, NMDA, GABA$_A$, and GABA$_B$ | 4       |
| $V_m$                                         | Values in mV | 1       |
| Temperature                                   | Values in °C | 1       |
| Inter-stimulus interval                        | Values in ms, NA (O) | 1       |
| Species                                       | Rats (1), mice (−1), or guinea pigs (O) | 1       |
| Sex                                           | Male (1), female (−1), or unstated (O) | 1       |
| Postnatal age                                  | Values in days | 1       |
| Slice region                                   | Hippocampal, entorhinal cortex, dorsal, ventral, and [medial or lateral] | 5       |
| Slice orientation                              | Coronal, sagittal, transverse, horizontal, longitudinal, and ‘magic cut’ | 6       |
| Slice thickness                                | Values in μm | 1       |
| Postsynaptic recording subregion               | Soma, dendrite, or axon initial segment | 1       |
| Ionic concentrations of intracellular and extracellular solutions | Intracellular and extracellular values in mols corrected for the ionic activity constants of solutions (Ca, Mg, Na, K, Cl, Cs, Br, Ba, $H_2PO_4$, $HPO_4$, $HCO_3$, gluconate, QX314, ATP, GTP, EGTA, OH, $SO_4$, phosphocreatine, acetate, methylsulfate, NMDG, tris, Ce$SO_4$) | 36      |
| Potency                                        | True vs false | 1       |

To assess model performance, we calculated training and prediction accuracies for the five synaptic parameters over all the data. Training accuracy measures how well the deep learning model predicts the synaptic parameters for a given pair of neuron types and a specific set of experimental covariates that were included in the training data. Prediction accuracy measures how well the deep learning model predicts the synaptic parameters for the data that were excluded from the training set. To assess this performance relative to the reliability of experimental measurements, we consider different experimental values (targets) recorded from the same nominal conditions (features). Those differences can be ascribed to unknown experimental factors, intrinsic biological variability, and random noise. We take such empirical ground-truth range as the gold standard to benchmark our model against. In these cases, we calculated the distance of each target value from their average, a measure of experimental fluctuation we call target variability (Fig. 4a and Suppl. Fig. 4b). The within-group sample size, summarizing the number of identical features with multiple measurements ($n$ values in Suppl. Fig. 4d) ranged from 2 to 10.

We compared the target variability with the training accuracy and prediction accuracy, i.e., the distance of model output from seen and unseen targets, respectively. The training and prediction accuracies of our predictive model were remarkably close to the target variability. Testing the predictive power of the model with the jackknife (leave-one-out) method, we found that the vast majority of unitary predictions fell within the 95% confidence interval of the targets, i.e., they were reliable (Fig. 4a, b). Specifically, this prediction reliability ranged from 90% for $\tau_r$ to 96% for $U$, with intermediate values for $g$ (91%), $\tau_d$ (94%), and $\tau_f$ (94%). By including all synaptic measurements (not just the unitary values), prediction reliability was reduced slightly to 88%–94% (Suppl. Fig. 4b–d). Additionally, comparing the relevant values to sparse estimates available for matching potential connections from a recent CA1 study$^{14}$ revealed no statistically significant difference for any of the five parameters (Suppl. Fig. 5a). The paired-pulse ratio of the models and the data before and after data normalization were also correlated, indicating both the TPM and the deep learning models had a reliable fit to the data (Suppl. Fig. 5b). Thus, the deep learning model quantitatively predicts the properties of synaptic signals for which experimental recordings are available within the margin of measurement accuracy.

**Connectivity matrix completion and synaptic data normalization.** Given its demonstrated performance on available data, the predictive model can confidentially estimate the synaptic parameters of yet uncharacterized potential connections based on the learned properties of neuronal types. The model can complete the synaptic electrophysiology matrix for all 3120 potential connections in the hippocampus and entorhinal cortex. Additionally, since the learned neuronal properties are all unique, the model also effectively disambiguates each potential connection: in other words, the predicted synaptic parameters for each pair of neuron types are also all unique. Notably, the deep learning model can infer synaptic parameters for every potential connection in any desired condition. Applying homogeneous conditions for all potential connections practically normalizes the inferences with...
respect to the covariates. This study primarily focuses on fast unitary synaptic properties in near-physiological (henceforth standard) conditions, namely AMPA and GABAA synapses of adult male rats in voltage-clamp at body temperature and with a pipette solution that does not disturb intracellular ionic concentrations (Methods). These so-derived synaptic signals reflected the training data and showed a wide range of amplitudes, kinetics, and ST-P across potential connections (Fig. 4c and Suppl. Movie 1). To explore regional differences within the hippocampal formation, we inspected the probability density distributions of all parameter values normalized using the min-max method (Suppl. Fig. 6a). Interestingly, the range of values in the entorhinal cortex is smaller than in the hippocampus. Moreover, the GABAergic and the glutamatergic synapses had overlapping distributions for g and U but not for the time constants (Suppl. Fig. 6b), suggesting that these synapse types have similar amplitudes but differ in decay kinetics and ST-P.

Open access to data and source codes. The normalized and completed synaptic data are broadly applicable to designing experiments in optimal conditions, testing hypotheses, constraining biologically plausible simulations of the entire entorhinal-hippocampal circuit23, and benchmarking machine learning algorithms. We provide five synaptic constants for each of 3120 connections in 32 different settings that include all binary
combinations of species (rat or mouse), sex (male or female), age (young or adult), recording method (voltage- or current-clamp), and temperature (room or body). For each parameter we make available the mean, standard deviation, and range over 100 training runs of the deep learning model (Fig. 5a). We also share all implemented tools for unhindered reuse with other datasets. The Synapse Modeling Utility, the preprocessing and analysis code in R, the machine learning library in Python, and the preprocessed machine learning-ready experimental data (2621 features-targets sets) are all freely available on Hippocampe.org/synapse (Fig. 5b).

Presynaptic and postsynaptic determinants of synaptic physiology. Deep learning-enabled full data normalization allowed us to compare the synaptic properties of all potential connections without the influence of confounding variables. Hereunder, we analyzed the normalized data. We reported trends in unnormalized data in an earlier paper11. To begin investigating how the presynaptic and postsynaptic identities combine to define synaptic dynamics, we asked two questions: (1) when a pair of neuron types forms a synapse, which synaptic properties (e.g., amplitude, duration, ST-P) does either side dominantly determine? (2) Does the answer differ for glutamatergic and GABAergic synapses? To answer these questions, we separated the glutamatergic and GABAergic synapses. In each pool, we created two groupings: one based on the presynaptic neuron types, and the other based on the postsynaptic ones. For example, the glutamatergic presynaptic grouping consisted of 38 groups, one for every glutamatergic presynaptic type; each group contains
Fig. 3 Synaptic modeling and deep learning. a Digitized trace (black line) from a CA2 Pyramidal cell to a CA2 Narrow-Arbor Basket PV+ cell. The corresponding 9-point reconstruction of each spike (red circles), based on initiation, peak, decay (filled circles), and 6 interpolations (hollow circles), are used to optimize the simulated signal (green). b The Tsodyks-Pawelzik-Markram (TPM) model describes synaptic amplitude, kinetics, and short-term plasticity in terms of utilization rate (\(u\)), activation (A), deactivation (D), and recovery (R) dynamics. State A represents the portion of activated synapses; state D corresponds to deactivated synapses that are still bound to neurotransmitter and therefore cannot be reactivated; R is the portion of synapses detached from neurotransmitter and ready to be reactivated. The kinetics of the transition from A to D is determined by the synaptic decay constant \(r_D\). The recovery rate is instead mostly determined by \(r_r\). In terms of postsynaptic ionotropic neurotransmitter receptors, the TPM model assumes that liganded-channel gap opens instantaneously. The portion of recovered synapses that are instantly activated after a synaptic event is indicated by \(u\) (lower case). During deactivation, the gate is closing while the neurotransmitter is still attached to the channel receptor. For a full recovery, the neurotransmitter needs to detach from the receptor for presynaptic reuptake. The time constant \(t_r\) measures that recovery speed. While \(r_r\) mainly defines synaptic decay, it can affect ST-P as well when presynaptic firing is very fast. TPM model combined with Ohm’s law can simulate voltage-clamp experiments (synaptic current, \(i_{syn}\)) by using experimentally measured reversal potential (\(E_{rev}\), junction potential (E), and holding potential (\(V_H\)). To simulate current-clamp experiments (synaptic potential, \(V_{syn}\)), we fed the \(i_{syn}\) to a simple membrane model (Resistor-Capacitor circuit), which depends on experimentally measured steady-state potential (\(V_m\), membrane capacitance (\(C_m\), and membrane time constant \(t_m\) that could be calculated knowing \(C_m\) and input conductance, \(g_m\)). A genetic algorithm yields the best-fitting values of 5 synaptic-specific parameters: conductance (g), single-exponential decay time constant \(\tau_d\), recovery time constant \(\tau_r\), facilitation time constant \(\tau_f\), and utilization ratio \(U\) (capital letter). c The predictive machine learning model of synaptic electrophysiology used a deep learning architecture with five hidden layers and error backpropagation. The input layer encoded the (typically fuzzy) presynaptic and postsynaptic neuron types (122 × 2 nodes) and all covariates (75 nodes). The output layer consisted of one node for each of the 5 synapse-specific parameters. Model training used the best-fitting TPM parameters corresponding to the available 2,621 reconstructed traces and matching covariates (corresponding to ~84% of 3,120 potential connections). The model outputs the five predicted synapse-specific parameters for any directional pair of neuron types and desired choice of covariates. We used random forest to complete the missing ST-P values (\(r_r\), \(r_D\), and \(U\)) in training data, and deep learning to predict synaptic properties of potential connections for which no data existed, including experimental conditions and potential connections that have never been studied before.

Principal covariate effects on synaptic properties. Next, we systematically investigated the influence of experimental covariates on synaptic parameters. Earlier research mainly checked the impact of experimental conditions on synaptic amplitude and kinetics of a limited number of neuron types. Our study also allowed the inclusion of ST-P parameters and systematically covered all potential connections of the hippocampal formation by changing one covariate at a time. All tested covariates had a statistically significant impact on synaptic parameters, but we only report here (Fig. 6b, c) those with a meaningful effect size (>10%) and emphasize the most substantive ones (>20%). Our results indicate that \(g\) increases more than two-fold and \(\tau_d\) decreases 30% when switching from voltage- to current-clamp, from male to female animals, and from gluconate-free to gluconate-containing intracellular solutions. While the change with recording modality agrees with previous studies, for example24, and we expected a difference by sex, the pronounced impact of gluconate in the pipette solution was surprising. Moreover, current-clamp (relative to voltage-clamp) and female animals (relative to male) also entailed notably higher \(r_D\) and lower \(r_r\), implying greater propensity towards synaptic depression. In contrast, the opposite trend, conducive to facilitation, was observed with gluconate. Shifting from rats to mice or from room to body temperature affected synaptic properties in the same direction, but to a more modest extent (10–20% effect size), as the male-to-female switch or intracellular gluconate addition, respectively. Reducing [CI], substantially increased short-term facilitation at GABAergic synapses, while more modestly slowing down synaptic kinetics which was unexpected based on25. Other covariates, including to our surprise age, did not affect the parameters substantially. Altogether, remarkably, only two types of variation, differing just in the change direction of \(r_r\) and \(r_D\), could explain the impact of all analyzed covariates irrespective of neurotransmitter type. This observation suggests an interdependence among synaptic parameters.

Synaptic amplitude predicts signal kinetics and the direction of short-term plasticity. Among both glutamatergic and GABAergic types, we noticed that synapses with high amplitude had fast kinetics and demonstrated depressing ST-P. Conversely, synapses with low amplitude had slower kinetics and were facilitating. To visualize these observations, we averaged the model parameters from the 30 synapses with the largest conductance and from the 30 with the smallest one among both glutamatergic and GABAergic groups. We then compared the responses of the four consensus models in standard conditions (Fig. 7a and Suppl. Movie 1). The high-amplitude models exhibited short-term depression and short signal duration (half-height width: 2.4 ms for glutamatergic and 3.8 ms for GABAergic), while the low-amplitude models demonstrated short-term facilitation and long signal duration (half-height width: 5.1 ms for glutamatergic and 6.2 ms for GABAergic). Considering all 3,120 connections revealed a significant negative correlation between \(g\) and \(\tau_d\) and between \(g\) and the paired-pulse ratio from baseline of the third synaptic event (\(AB_2A_1\)), but a positive correlation between \(g\) and \(U\), suggesting that high-amplitude synapses have higher resource utilization (Fig. 7b). Facilitation and depression partly depend on interstimulus intervals (ISI) and the measure of ST-P. Testing ST-P at 20 ms ISI and considering \(AB_2A_1\), the majority (>90%) of synapses with an amplitude below 0.5 nS facilitated, irrespective of neurotransmitter, while most synapses above 2 nS (glutamatergic) or 3 nS (GABAergic) depressed (Fig. 7c, left). Although the second synaptic events (\(AB_2A_1\)) tended towards facilitation relative to subsequent signals (e.g., \(AB_3A_1\)), all ST-P measures consistently transitioned from facilitation to depression as a function of conductance (Fig. 7c, right). Moreover, \(r_d\) and \(r_r\) were...
negatively correlated ($R_{\text{glu}} = -0.4$, $R_{\text{GABA}} = -0.1$, $p < 0.05$), indicating that synapses needing a long time to recover their resources tend to reduce their synaptic utilization rate rapidly. Altogether, these analyses suggest that higher synaptic amplitudes predict faster kinetics and a tendency towards depression over facilitation, reflecting coordinated differences in $\tau_d$ and $U$ as well as interdependence of $\tau_f$ and $\tau_r$.

Presynaptic and postsynaptic molecular expression as a biomarker of short-term plasticity. It is a widespread practice to
study synapses based on molecular expression. Chemical biomarkers were not directly among the training features of our predictive synapse models, but were used for mapping mined signals to potential connections. Since the normalized inferences are not fuzzy, we employed Hippocampome.org to query neuron types expressing different markers and analyzed differences in synaptic properties among neuron types containing (+) or lacking (−) each molecule. Since marker expression can be localized to the presynaptic terminals or postsynaptic dendrites and soma, we studied the presynaptic and the postsynaptic groups separately (Fig. 8).

AB3:A1 as a measure of ST-P and using a 20 ms ISI, we identified two classes of presynaptic markers that respectively predicted synaptic facilitation and depression. Specifically, presynaptic calbindin (CB), cholecystokinin (CCK), and neuropeptide-Y (NPY) expression correlated with facilitation (larger AB3:A1 values). In contrast, calretinin (CR), parvalbumin (PV), and somatostatin (SOM) correlated with depression (smaller AB3:A1 values). The relations of these markers with changes in synaptic amplitude and kinetics were not always statistically significant but generally followed the trends revealed in the previous section. Namely, presynaptic expressions predicting short-term facilitation typically demonstrated lower signal amplitudes and slower kinetics and vice versa for those predicting short-term depression. Cannabinoid receptor 1 (CB1) expression can be localized both on presynaptic and postsynaptic sides of a synapse. Since the presynaptic effects were similar to CCK, we only illustrated the postsynaptic effects. Among the postsynaptic markers, CB1 and serotonin receptor 3 (5HT-3) predicted lower amplitudes and a tendency towards facilitation. Interestingly, CB1 changed g and AB3:A1 of GABAergic synapses more than of glutamatergic synapses, on average.

**Correlations between neuronal morphology and synaptic parameters.** In GABAergic neurons of both hippocampal area CA1 and visual cortex, the kinetics of spontaneous synaptic inputs vary depending on the specific axonal targeting of that same postsynaptic neuron. We tested similar interactions between input synaptic properties and output axonal patterns throughout the hippocampal formation, not only considering unitary synaptic kinetics, but also conductance and ST-P (Suppl. Figure 7). Among GABAergic synapses in CA1, we found significant differences in g, τd, τf, and U, indicating that input synaptic duration, as well as amplitude and facilitation, vary by output axonal targeting (Fig. S7a). Extending the study to other hippocampal regions revealed significant differences in...

---

**Fig. 5 Comprehensive open access to data and tools.**

- **a.** The full set of normalized synaptic data for the entire entorhinal-hippocampal circuit in 32 different conditions (any combination of rat/mouse, male/female, young/adult, body/room temperature, and voltage-/current-clamp), the reconstructed synaptic traces with original references and annotated metadata, and the machine learning training data are all freely available at Hippocampome.org/synapse. For each synaptic parameter of every potential connection, we supply the mean, standard deviation, and range of 100 deep-learning model predictions.

- **b.** Our high-performance synapse modeling tool (Hippocampome.org/SynapseModelingTools) is equipped with a Trace Reconstructor and trace correction algorithm.
τ_d and τ_f among GABAergic synapses in CA3, and in τ_r in DG and CA2. Glutamatergic synapses generally demonstrated fewer significant differences. Visualizing consensus traces (Suppl. Figure 7b) and synaptometrics differences (Suppl. Fig. 7c) confirmed these patterns.

In the visual cortex, connection probability correlates with synaptic strength.32 Hippocampome.org calculates the probabilities of connections and the average synaptic distance from the presynaptic and postsynaptic soma based on the layerspecific linear densities of the corresponding axons and dendrites.33 Synaptic conductance had a weak but statistically significant positive correlation with the connection probability (R_GABA = 0.27, R_Glu = 0.19, p < 0.05). Consistent with dendritic filtering, we found a significantly negative correlation between g
and the synaptic distance from the postsynaptic soma ($R_{GABA} = -0.13, R_{Glu} = -0.06, p < 0.05$).

**Discussion**

We digitized, reconstructed, and compiled a comprehensive dataset of 2621 synaptic signals recorded from the rodent hippocampus and entorhinal cortex, and mapped each to respective covariates and potential connections. Through computational modeling and machine learning, we normalized and completed all synaptic physiology data to predict the amplitude, kinetics, and ST-P of the 3120 potential connections of the hippocampal formation. For each potential connection, we freely released via Hippocampome.org the complete set of 5 synaptic parameters in 32 different experimental settings with all annotated experimental data, plus analysis and modeling software source code. We identified the major determinants of unitary synaptic physiology and discovered new correlations among synaptic properties, molecular expression, and neuronal morphology.

Broad diversity in experimental settings causes extreme variability in synaptic electrophysiological recording. Combined with inherent measurement noise, this makes identifying causal relations among variables considerably challenging. To our knowledge, our application of deep learning provides a suitable solution to these issues. Testing the deep learning model with unseen data demonstrated that the predictions are valid within experimental accuracy. Applying uniform experimental conditions (voltage-clamp at body temperature in male rats with specific experimental conditions) to all potential connections effectively normalized data. In that scenario, the only differences in synaptic parameters are due to the presynaptic and postsynaptic neuron identities. At the same time, changing the chosen experimental condition, such as switching from male to female animals, allows the systematic investigation of every covariate effect.

Furthermore, our deep learning solution yields two notable data augmentation benefits. First, it fills in missing data by matrix completion harnessing the learned axonal and dendritic properties of the corresponding neuron types. In simple terms, if the predictive model learns synaptic features from neuron type x to neuron type y, and from type w to type z, it can then infer the features from x to z and from w to y based on the axonal properties of x and w and the dendritic properties of y and z. In reality, the known features utilized in training are more numerous than the set of missing data. For comparison, an earlier study measured the synaptic physiology of 10% of potential connections in CA1 to extrapolate the properties of the remaining 90% based on marker profiles. In contrast, our experimental dataset covered most potential connections across the entire hippocampal formation, with missing values ranging from 16.3% for conductance to 38.5% for ST-P. Singular value decomposition (SVD) may robustly complete matrices with up to 50% of missing values, but deep learning typically outperforms SVD in this process.

The second beneficial effect of our machine learning approach is that it leverages data redundancy to disambiguate the mappings of individual signals to multiple potential connections. Consider for instance an experimental recording mapped to potential connections A or B and a different recording mapped to potential connections B or C; the deep learning model utilizes the two constraints on B to predict a unique set of synaptic parameter values distinct from those of A and C. Indeed, the inferred values were all different for the 3120 pairs of hippocampal neuron types, indicating that the training data was sufficient to completely resolve degenerate mappings.

The method we introduced is highly flexible and can be adapted to include pathological conditions. For example, the set of features could be expanded by including a descriptor to distinguish the epileptic state. Then, mining and modeling available epileptic data, would allow the inference of normalized and completed synaptic properties in epilepsy. At this time, we have mined only the control condition of experiments. However, our freely accessible code and data allow any interested researcher to add and normalize their recordings as needed. The more data are pooled together, the more generalizable the resulting predictions will be.

Before delving into the implications of the results, we should note that our observations are circumscribed to the somatic impact of synaptic events. To investigate local postsynaptic mechanism likely requires compartmental modeling of dendritic morphologies, or optoelectrophysiological techniques. We also did not consider long-term plasticity and slow conductances ($GABA_b$ or NMDA receptors; but see Suppl. Note 1) since the required experimental data for most pairs of neuron types remain sparse. Our regression models likewise excluded stimulation strength of evoked events because this detail is seldom reported in publications. Additionally, most published traces had a constant ISI and a single recovery event. Future experimental designs including variable ISIs and multiple recovery events would allow more accurate estimation of $\tau_r$ and $\tau_f$. Since homeostatic plasticity may change synaptic strength in vitro, we suggest reporting systematically the time elapsed from slice preparation until the actual recording. It is also important to acknowledge that the chosen TPM formalism is a fairly simple analytical model. While it performs satisfactorily for the majority of synapse types in the hippocampal formation, more complex models accounting for rise time constant, calcium concentration, and stochasticity may be better suited for extremely facilitating synapses (see Suppl. Note 2).

The synapses of the entorhinal-hippocampal network communicate through a broad continuum of signal amplitudes. Yet, the sets of neurotransmitters and receptors employed by this circuit are limited, raising a question: does variation in synaptic conductance interact with resource utilization and recovery to affect kinetics and ST-P? Unnormalized unitary data suggest that decay kinetics are faster for strong GABAergic synapses than for weak ones. Additionally, one study on three synapse types suggests that the ST-P of stronger synapses is depressing, and the ST-P of weaker synapses is facilitating. Indeed, analyzing all potential connections of the hippocampal formation revealed a
Fig. 7 Synaptic conductance entails kinetics and short-term plasticity. 

**a** We averaged the synaptic parameters of the 30 potential connections with the largest conductance (high-amplitude) and of the 30 with the smallest one (low-amplitude) to compare their simulated consensus signals. The low- and high-amplitude groups exhibited respectively short-term facilitation and depression (ISI = 20 ms, $V_h = -35$ mV, $GABA_{E_{rev}} = -70$ mV, and AMPA $E_{rev} = 0$ mV). $A_1$ is the synaptic amplitude of the first event; $AB_3$ is the amplitude from the baseline of the third event. 

**b** Considering all connections, synaptic amplitude correlates with kinetics and short-term plasticity (ST-P). The negative correlation between $g$ and $\tau_d$ (left) indicates that synapses with large conductance tend to have faster signal decays. The positive correlation between $g$ and $U$ (middle) suggests that high-amplitude synapses have higher resource utilization. The negative correlation between $g$ and the ratio from the baseline of the third event, $AB_3:A_1$ (right) shows that high-amplitude synapses depress more. 

**c** The ratio between facilitation ($AB_3:A_1 > 1$) and depression ($AB_3:A_1 \leq 1$) decreases as a function of conductance with a transition from mainly facilitating to mainly depressing between 1 nS (GABAergic) and 1.7 nS (glutamatergic) for 20 ms ISI. 

**d** Fitting the dependence on $g$ of the ratio from the baseline of the $x^{th}$ event, $AB_x:A_1$, with an inverse first-order polynomial function reveals that earlier synaptic events (e.g., $AB_2:A_1$) tend to facilitate while later ones (e.g., $AB_5:A_1$) tend to depress. The trend from facilitation to depression with increasing conductance is robust to ST-P measure.
negative correlation of \( g \) with both \( \tau_d \) and \( AB_2:A_1 \). Moreover, we found a positive correlation between \( g \) and \( U \), consistent with the TPM model (Eq. 18 in Methods). Since \( U \) quantifies the utilization increment, these results suggest high-amplitude synapses depress more easily because of resource exhaustion.

The TPM model accounts for resource utilization and recovery. When \( \tau_r \) is small, the resource recovery pace is fast. When \( \tau_f \) is large, resource utilization remains prolonged. Therefore, the opposite dependence of \( \tau_f \) and \( \tau_U \) on covariates indicates that when resource recovery pace is fast, resource spending is prolonged. Furthermore, their higher negative correlation in glutamatergic synapses relative to GABAergic ones suggests that resource utilization is subject to tighter control in the former than in the latter. Overall, the effects of covariates on synaptic parameters revealed only two distinct patterns that differed exclusively in the change direction of \( \tau_f \) and \( \tau_r \). The correlation among synaptic parameters could explain the mere simplicity of these observations. Covariates increasing \( g \) will also increase \( U \) and decrease \( \tau_d \). The only remaining freedom is in \( \tau_f \) and \( \tau_r \), which always change in opposite directions. This suggests that covariates affect a small set of latent variables. See Suppl. Note 3 for further discussions.

For equivalent experimental conditions and irrespective of the neurotransmitter, female animals had, relative to males, multiplicatively larger unitary synaptic conductance, significantly faster kinetics, and greater tendency towards short-term depression than facilitation. It is tempting to speculate a link to chronic exposure to neurosteroids and endocannabinoids, which increase the amplitudes of glutamatergic and GABAergic synapses, respectively, in females. We observed similar changes in synaptic parameters when switching from voltage-clamp to current-clamp. This could be due to the activation of voltage-gated ion channels in current-clamp or the reduction of passive filtering during parametric fitting that brings the estimations closer to the local dendritic event. We also found qualitatively parallel differences between species, with significantly larger synaptic conductance in mice compared to rats. Notwithstanding the high statistical sensitivity of our study, however, the phenomenological disparity across rodents was practically negligible (see Suppl. Note 4 for further considerations).

When added to the patch-clamp intracellular solution, the common food additive potassium gluconate (E577) changes the reversal potential of GABA\(_A\) channels, blocks ion channels involved in subthreshold membrane physiology, and alters firing patterns in hippocampal neurons. However, the impact of intracellular gluconate on unitary synaptic signaling has never been studied systematically. We found intracellular gluconate to be one of the most potent synaptic enhancers. With gluconate in the recording pipette, synaptic amplitudes were a fold-factor larger, kinetics were faster, and short-term plasticity shifted from depression to facilitation (smaller \( \tau_d \) and larger \( \tau_f \)). The increment of synaptic amplitude could be explained by blockage of the subthreshold channels, which reduces shunting and increases input resistance. The reduction of short-term depression may be due to the role of gluconate as an energy source that facilitates
resource recovery. As a comparison, the effect of gluconate on synaptic parameters was a full order of magnitude larger than the changes observed in the same direction when shifting from room temperature to body temperature.

Our data analysis suggests that the presynaptic side of glutamatergic, and the postsynaptic side of GABAergic neurons, have a relatively higher impact on synaptic properties. For GABAergic synapses, this finding could be explained by the selective targeting of Axo-axonic and Interneuron-Specific interneurons. At the same time, each neuron type in Hippocampome.org is linked to known molecular biomarkers expressed either in the axons (e.g., calcium-binding proteins and neuropeptides) or in the dendrites (e.g., neurotransmitter receptors). Among calcium-binding proteins, calbindin was a biomarker of facilitating synapses while calretinin and parvalbumin of depressing ones. Among neurotransmitters, cannabinoid receptor 1 and serotonin-gated ionotropic receptors, cannabinoid receptor 1 and serotonin-gated ionotropic channels altered synaptic properties similarly. While this result is consistent with their pattern of co-expression in cortical neurons, CB1. See Suppl. Note 5 for further discussions.

Biophysical synaptic model and parametric fitting. To facilitate comparison between current and voltage recordings, we reduced the signals to modality independent synaptic constants utilizing a specific version of the Tsodyks, Pawelzik, and Markram (TPM) model. The TPM model formulates a relationship between synaptic conductance (g), deactivation time constant (τd), recovery time constant (τr), facilitation time constant (τf), and the utilization ratio (U) of synaptic resources in one set of ordinary differential equations. Calculating synaptic currents (I_syn) with an Ohmic model for ion channels (Fig. 3b) requires the reversal potential (E_rev) and the postsynaptic membrane potential (V_m). E_rev is experimentally measurable or can be accurately estimated from the ionic composition of bath and pipette solutions, temperature, and permeability of ion channels to different ions. We assigned kinetically fast synaptic responses to be mediated by calcium-impermeable AMPA or GABA_A channels, unless otherwise stated in the original publications. Because I_syn is recorded in voltage-clamp experiments, we calculated V_m by correcting holding potential (V_h) for liquid junction potential (E_j) as previously described.

Using the TPM model, we can analytically simulate the amplitude, kinetics, and ST-P of I_syn. We numerically derived synaptic potentials (V_m) via feeding the simulated I_syn to a resistor-capacitor circuit (RC) model of neuronal membrane, from which we estimated V_m as the evolution of V_m over time. We used the OpenModelica solver via SciPy for numerical integration. This model depends on three experimentally measurable parameters: the membrane time constant (τ_m), membrane capacitance (C_m), and the initial value of V_m. Since I_syn is recorded in current-clamp experiments, we corrected resting or steady-state membrane potential for E_j to estimate the initial value of V_m. We used τ_m and C_m values when reported in the original study; otherwise, we utilized the values reported by Hippocampome.org for a matching synaptic model. The available temperature, recording method, and solutions were optimized during parametric fitting. Only for 23% (603:2621) of the signals at least one of the τ_m and C_m values was found through optimization.

We found the optimal g, τ_d, τ_r, and U values for each experimentally recorded synaptic signal by fitting TPM model simulations to the reconstructed data points. We created a high-performance and user-friendly Python simulator, the Synapse Modeling Utility, to aid in parametric fitting. Optimization was performed by an implementation of the SciPy toolbox genetic algorithm, a bound-constrained global optimizer. As the objective function, we chose the mean soft L1 squared error, i.e.,

\[ error = \frac{1}{n} \sum_{i=1}^{n} \left( \sqrt{1 + \left( \frac{\text{signal} - \text{simulation}}{\text{signal}} \right)^2} - 1 \right) \]

where n is the number of reconstructed data points. We assigned the fitting error associated with the first synaptic event twice the weight relative to all other events, and the 6 interpolated data points half the weight of the initiation, peak, and decay points divided by the following bound constraints: 1 < τ_d < 3000 ms, 1 < τ_r < 3000 ms, and 0.001 < U < 1. Optimization was stopped when the difference of fitting error between successive fits yielded a change of less than 0.001.

If more than one stimulation frequency was available for a given experiment, we pooled data prior to optimization to ensure the estimated parameters are more generalizable to different frequencies. We then re-expanded the data after optimization to match each of the original traces.

The Synapse Modeling Utility also implemented a correction for slow processes when the ISIs varied (see also Supplementary Note 1). In the absence of a slower process, the signal should gradually decay to a baseline. Provided slower processes do not drastically affect the first simulated event such that τ_d estimation is accurate, the recorded and simulated signals should be most similar at the initiation points of the synaptic events. We defined the correction amounts at initiation points of two successive synaptic events as the amplitude differences (Suppl. Fig. 1b). We then calculated the correction values for data points in between two initiation points by the triangulation method. Since τ_d values greater than 700 ms were excluded from subsequent analysis. For signals that only reported amplitude and kinetics, we set the missing τ_d values to the median of the unitary GABAergic and glutamatergic responses as appropriate.

Fitting a single synaptic signal using our Synapse Modeling Utility required seconds to minutes, while with a pipeline built using the Neuron Simulation Environment required hours. We optimized each trace at least 30 times and averaged the best 15 fits. The relative inter-trial variability was <0.001.
Machine learning design and implementation. We employed machine learning to infer the five parameters of the TPM model (the targets) based on a set of features. In addition to the presynaptic and postsynaptic intensities, the TPM model parameters had units (i.e., the synaptic conductance in pS and cm⁻²). Specifically, the set of features consisted of 319 one-hot encoded and numerical values (Table 1): 122 features encoded presynaptic neuron types, 122 postsynaptic neuron types, and the remaining 75 features encoded the experimental covariates. For instance, three columns encoded four stimulation methods (evoked, spontaneous, forward, and backward), and the remaining 75 features encoded all the experimental covariates (e.g., cell types). Each row was associated with a presynaptic spike, and all rows captured a single trial. Consequently, the dataset naturally grows backward by backpropagation. For each event, we inferred the target synaptic amplitude or potency, which differs in the different datasets: if failed events are not reported, we estimated it based on weight, diet, species, and strain. The remaining features encoded whether the target signal represented amplitude or potency, which differs in the averaging method: if failed events are excluded from the signal average, the peak quantifies synaptic potency rather than amplitude. For algorithm training, when the failure rate of the first synaptic event was reported, we added an additional pseudo-signal by converting amplitude to/from potency, resulting in a different g value as a target and a different potency value as a feature. We normalized features with the MaxAbsScaler function of Scikit-learn to preserve the sparsity of the feature matrix. Moreover, we normalized the targets with the MinMaxScaler method to allow usage of sigmoid activation functions in the deep learning output layer.

We implemented the machine learning pipeline in PyTorch and Python. First, we trained random forests using the Scikit-learn toolbox on two Xeon-E5 v3 CPUs to infer missing values of ST-P. The random forest model is a retrained random forest model using the Scikit-learn package on two Xeon-E5 v3 CPUs to infer missing values of ST-P. The random forest model is a refined series of linear regressions that correct the predictions in every step towards the final output. This algorithm is fully automated, has only one tuning parameter, and is very robust. Specifically, whenever a signal lacked estimates for 

\[
\text{prediction} = \frac{\sum_{i=1}^{n} \text{predictions}_i}{n}
\]

and 

\[
\text{target} = \frac{\sum_{i=1}^{n} \text{targets}_i}{n}
\]

we set the ISL as a feature, to zero in those cases. The random forest model, in contrast to deep learning, could learn to predict missing values for 

\[
\text{prediction} = \text{prediction} + \text{ISL}
\]

and 

\[
\text{target} = \text{target} + \text{ISL}
\]

for a different potency value as a feature. We normalized features with the MaxAbsScaler function of Scikit-learn to preserve the sparsity of the feature matrix. Moreover, we normalized the targets with the MinMaxScaler method to allow usage of sigmoid activation functions in the deep learning output layer.

We implemented the machine learning pipeline in Jupyter and Python. We first trained random forests using the Scikit-learn toolbox on two Xeon-E5 v3 CPUs to infer missing values of ST-P. The random forest model is a refined series of linear regressions that correct the predictions in every step towards the final output. This algorithm is fully automated, has only one tuning parameter, and is very robust. Specifically, whenever a signal lacked estimates for 

\[
\text{prediction} = \frac{\sum_{i=1}^{n} \text{predictions}_i}{n}
\]

and 

\[
\text{target} = \frac{\sum_{i=1}^{n} \text{targets}_i}{n}
\]

we set the ISL as a feature, to zero in those cases. The random forest model, in contrast to deep learning, could learn to predict missing values for 

\[
\text{prediction} = \text{prediction} + \text{ISL}
\]

and 

\[
\text{target} = \text{target} + \text{ISL}
\]

for a different potency value as a feature. We normalized features with the MaxAbsScaler function of Scikit-learn to preserve the sparsity of the feature matrix. Moreover, we normalized the targets with the MinMaxScaler method to allow usage of sigmoid activation functions in the deep learning output layer.

Machine learning model validation. We computed training accuracy as the average SMAPE (symmetric mean absolute percentage error) of the model output from the training data. In contrast, prediction accuracy is the SMAPE of the model output from the unseen data. We monitored the prediction accuracy of the model after each training epoch with k-fold cross-validation with k = 4. We trained four models on four separate training runs, each using three-quarters of the data, used the remaining one-fourth of data to measure prediction error, and averaged the results over the four runs. We assessed our model accuracy with the jackknife method (Fig. 4a, b and Suppl. Fig. 4): we trained n = 2621 models with n − 1 data points and assessed the prediction error of the model for one set aside data point.

Our dataset had more than one data point for most potential connections (Fig. 2, and Suppl. Fig. 2e, f). In certain cases, these data points had identical values. For instance, we selected identical ionic concentrations for physiological solutions across all synapses and calculated E_rev accordingly. We set no NMDA or GABA_B contamination for the features. Using the trained deep learning models, we inferred unitary synaptic parameters for each potential connection always verifying that the predicted values remained within the upper and lower boundaries of the training set to avoid erroneous extrapolations. We chose unitary postsynaptic currents recorded from adult male rat slices kept in artificial cerebrospinal fluid at body temperature while using whole-cell patch pipettes devoid of high [Cl]i, or [glucose], solutions as a standard condition for model inferences. When analyzing covariates, we changed one feature at a time to infer the corresponding synaptic parameter. We also generated the references for different conditions, i.e., rat vs mice, male vs female, P14 (adolescent) vs P56 (adult), room temperature vs body (32 °C) temperatures, and voltage-clamp vs current-clamp recording methods.

Data normalization. The training features were highly heterogeneous and typically mapped to multiple presynaptic and/or postsynaptic neural types (fuzzy or ambiguous mapping). Nevertheless, a trained model can predict targets (synaptic parameters) for an arbitrary set of features. We inferred values for the unambiguous (proper) mapping of all 3120 potential connections in the entorhinal-hippocampal network. In other words, each inference feature was mapped to one presynaptic neuron type and one postsynaptic neuron type. We also set all other features except the presynaptic and postsynaptic neuron types to identical values. For instance, we selected identical ionic concentrations for physiological solutions across all synapses and calculated E_rev accordingly. We set no NMDA or GABA_B contamination for the features. Using the trained deep learning models, we inferred unitary synaptic parameters for each potential connection always verifying that the predicted values remained within the upper and lower boundaries of the training set to avoid erroneous extrapolations. We chose unitary postsynaptic currents recorded from adult male rat slices kept in artificial cerebrospinal fluid at body temperature while using whole-cell patch pipettes devoid of high [Cl]i, or [glucose], solutions as a standard condition for model inferences. When analyzing covariates, we changed one feature at a time to infer the corresponding synaptic parameter. We also generated the references for different conditions, i.e., rat vs mice, male vs female, P14 (adolescent) vs P56 (adult), room temperature vs body (32 °C) temperatures, and voltage-clamp vs current-clamp recording methods.

Statistics and reproducibility. To compare synapses, we either analyzed the TPM model parameters directly or simulated each synapse separately and measured different synaptometrics. The paired-pulse ratio is the measure of ST-P, which requires the estimation of amplitude. For the first synaptic event, the baseline crosses the initiation points, but for later events, the overlap of initiation points with the baseline depends on the ISI and f values. If the amplitude is measured at baseline, we calculated the paired-pulse ratio; otherwise, if the amplitude is measured at the initiation point, we calculated the A1 term. For example, A1 is the amplitude of the first synaptic event, and A2B is the amplitude of the baseline of the third synaptic event (Figs. 3 and 7). Thus, A1/A2 represents the percentage change of the ith event from baseline, which assesses the weakening of the synaptic activation (see derivation of TPM model section for mathematical clarification).

We compared groups with paired or unpaired Wilcoxon’s test as appropriate. We corrected all p-values for multiple comparisons using False Discovery Rate and selected 0.05 as the significance threshold. We corrected coefficients of variation (CVs) for sample sizes and used the bootstrapping method to find the confidence intervals. We used the Pearson method to compute correlations. The p-value of correlations is calculated using t distribution table and ggpubr package in R.

As a measure of central tendency, we defined trimmed-mean as a mean value in which 2.5% of outliers on both extremes are excluded. The interquartile range is the measure of spread. Since the synaptic parameters have different units, we used symmetric percentage distance

\[
\text{SPD} = \frac{\max(\text{change} - \text{reference}, 0) + \max(\text{reference} - \text{change}, 0)}{2}
\]

as a measure of change between two data points that is dimensionless and unbiased. We simply use the term percentage to refer to SPD in the Results and Discussion of this paper. Specifically, for covariates analysis (Fig. 6b), we computed the trimmed-mean of SPDs of reference vs change of each potential connection:

\[
\text{SPD} = \frac{\max(\text{change} - \text{reference}, 0) + \max(\text{reference} - \text{change}, 0)}{2}
\]

Then, compared the differences with paired Wilcoxon’s test. For morphology and marker analysis (Fig. 8b, c), we used the unpaired Wilcoxon’s test.
We converted the Wilcoxon’s estimate, a robust measure of the difference between groups, and 95% confidence intervals to SPD by multiplying these values by

\[
200 \text{median}(-\delta) + \text{median}(\delta)
\]

where (+) refers to the group that expressed the marker and (−) for the group that did not.

**Derivation of a simplified Tsodyks, Pawelzik and Markram synapse model.**

Over the last 50 years, a large body of phenomenological synaptic plasticity models has been theorized\(^1\). One of the better-established models is that of Tsodyks, Pawelzik, and Markram (TPM)\(^1\). In this work, we adapted a simplified version of the TPM model\(^1\) and further streamlined the analytical solutions.

**Ordinary differential equations describing synaptic temporal dynamics.** Short-term synaptic plasticity depends on the availability and utilization of synaptic resources (Fig. 3b), including the number of readily releasable synaptic vesicles and the concentration of calcium. Short-term synaptic facilitation begins with an increase of calcium ions within the presynaptic terminal resulting in an increase in synaptic resource utilization. Equation 6 formulates the utilization dynamics:

\[
du \frac{dt}{dt} = -\frac{u}{\tau_{u}} + U \cdot (1 - u_{-}) \cdot \delta(\Delta t) \quad \text{with} \quad \Delta t = t - t_{i}
\]

where \(u\) is the fractional degree of synaptic utilization at any moment \(t\), \(u_{-}\) indicates the value of \(u\) just before the synaptic event time \(t_{i}\), \(U\) determines the increment proportion (between 0 and 1) with each presynaptic spike, and \(\delta\) is Dirac’s delta function. Since \((1 - u)\) quantifies unutilized resources and synapses cannot use more than all the resources available to them, \(U \cdot (1 - u_{-})\) determines \(u\) increment after each synaptic event. Whenever a synapse is not being stimulated, synaptic utilization exponentially decays to zero with the facilitation decay time constant \(\tau_{u}\).

Synaptic depression is due to the depletion of available synaptic resources. These resources can be partitioned into three portions, representing respectively the activated (A), deactivated (D), and recovered (R) states. After each presynaptic spike, an instantaneous shift occurs from recovered to activated state. The amount of shift is determined by \(u\). The active resources then decay to the deactivated state by the decay time constant \(\tau_{D}\). Since synaptic resources are limited, the more resources stay in the deactivated state, the more a synapse is depressed. In the TPM model, synaptic resources exponentially recover from depression with the recovery time constant \(\tau_{r}\). This process can be formulated by the following set of equations:

\[
\frac{dR}{dt} = \frac{D}{\tau_{r}} - u_{+}A_{-}R_{-}\delta(\Delta t)
\]

\[
\frac{dA}{dt} = -\frac{A}{\tau_{A}} + u_{+}R_{-}\delta(\Delta t)
\]

\[
\frac{dD}{dt} = \frac{A}{\tau_{D}} - \frac{A}{\tau_{r}} + u_{+}R_{-}\delta(\Delta t)
\]

where \(u_{+}\) is the value of \(u\) just after synaptic event time, which can be determined using Eq. 6. \(R_{-}\) is the value of \(R\) just before the synaptic event, which is determined by Eq. 7. The product \(u_{+}R_{-}\) represents the fraction of the synaptic resources being utilized after each synaptic event. This proportion is added to the already active resources (\(A\)) and taken from the readily usable resources (\(R\)). Then, the change in \(D\) at any moment is the difference between resources deactivating (\(D/\tau_{D}\)) and resources recovering (\(A/\tau_{r}\)).

A simplified version of the four-state TPM model\(^1\) eliminated Eq. 9 which is possible since the total amount of synaptic resources is fixed:

\[
R + A + D = 1
\]

Substituting \(D = 1 - R - A\), the four-state TPM model can be reduced to the following three-state model:

\[
\frac{d\delta}{dt} = -\frac{\delta}{\tau} + U \cdot (1 - u_{-})\delta(\Delta t)
\]

\[
\frac{d\delta}{dt} = \frac{1 - \delta}{\tau_{D}} - u_{+}A_{-}R_{-}\delta(\Delta t)
\]

\[
\frac{d\delta}{dt} = -\frac{\delta}{\tau} + u_{+}R_{-}\delta(\Delta t)
\]

**Analytical solution of the model.** This three-state model can be solved using the technique of exact integration\(^1\). If \(\frac{d}{dt}\) is a time varying function of \(S(t)\),

\[
\frac{df(t)}{dt} = \frac{1}{\tau}\left(-f + S(t)\right)
\]

**Summary of the analytical solution.** Since \(A(t)\) is independent of the rest of the equations, the simulation of synaptic amplitude after each synaptic event only requires the calculation of Eq. 14. When a synaptic event occurs, the value of each of the states should be calculated just before the synaptic event.

\[
\begin{align*}
A(t) & = A_{+} - A_{-} \quad \text{if} \quad t \neq t_{i} \\
A(t) & = A_{+} \quad \text{if} \quad t = t_{i}
\end{align*}
\]

Similarly, the solution for \(R(t)\) is:

\[
\begin{align*}
R(t) & = R_{+} - R_{-} \quad \text{if} \quad t \neq t_{i} \\
R(t) & = R_{+} \quad \text{if} \quad t = t_{i}
\end{align*}
\]

The solution for \(D(t)\) is:

\[
D(t) = D_{+} - D_{-} \quad \text{if} \quad t \neq t_{i} \\
D(t) = D_{+} \quad \text{if} \quad t = t_{i}
\]

**The first synaptic event.** Ohm’s law is used to calculate the synaptic currents (\(I_{syn}\)):

\[
I_{syn} = g \times (V_{m} - E_{rev})
\]

In the TPM model, \(I_{syn}\) is calculated with the following equation:

\[
I_{syn} = g_{optimization} \times A \times (V_{m} - E_{rev})
\]

Therefore,

\[
g = g_{optimization} \times A
\]
Before any synaptic event, all resources are readily usable, and there is no utilization and activation. Therefore,
\[
\begin{align*}
    u_t &= 0 \\
    A_t &= 0 \\
    R_t &= 1
\end{align*}
\]
For the first synaptic event, the value of \( A \) is easily calculable.
\[
\begin{align*}
    u_t &= u_0 + U (1 - u_0) = U \\
    A_t &= A_0 + u_t R_t = u_t
\end{align*}
\]
Therefore,
\[
\begin{align*}
    g &= g_{\text{optimization}} \times A_t = g_{\text{optimization}} \times U
\end{align*}
\]

**Distinction between short-term plasticity measures.** The distinction between the paired-pulse ratio (PPR1) and paired-pulse ratio from the baseline (\( AB : A_t \)) is formulated with the following equations:
\[
\begin{align*}
    \text{PPR}_{1} &= \frac{A_t - A_{t-1}}{A_t - A_{t-1}} = \frac{u_t R_t}{u_{t-1} R_{t-1}} \\
    \text{AB} : A_t &= A_t \frac{1}{A_{t-1}}
\end{align*}
\]
These equations indicate that the \( AB : A_t \) measures the evolution of a state but \( \text{PPR}_{1} \) the evolution of \( u_t R_t \).

**Convergence of numerical and analytical solutions.** We implemented the numerical and analytical solutions in the NEURON simulation environment and compared them to the original four-state model to confirm the convergence of all the formalisms (Suppl. Fig. 9). The simulation files are available to download from the ModelDB portal (Accession: 266934).

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**
Supplementary Data 1 provides the source data underlying Figs. 6a and 8. All other data are released on Hippocampome.org/synapse. The underlying experimental measurements come from Hippocampome.org/synaptome as described in. Supplementary Data 2 provides the list of all 160 articles reporting those measurements.

**Code availability**
The Synapse Modeling Utility developed in this work is available at https://github.com/k1moradi/SynapseModelingUtility, https://doi.org/10.5281/zenodo.6385860, and https://hippocampome.org/general/synapse_modeling/Modeling.rar. The Machine Learning Library is available at https://github.com/k1moradi/MachineLearningSynapsePhysiology, https://doi.org/10.5281/zenodo.6385648, and https://hippocampome.org/general/synapse_modeling/MachineLearningSynapsePhysiology.zip.

Received: 18 June 2021; Accepted: 30 March 2022; Published online: 05 May 2022

**References**
1. Hafting, T., Fyhn, M., Molden, S., Moser, M. B. & Moser, E. I. Microstructure of a spatial map in the entorhinal cortex. Nature 436, 801–806 (2005).
2. O’Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res. 34, 171–175 (1971).
3. DeFelipe, J. From the connectome to the synapome: an epic love story. Science 330, 1198–1201 (2010).
4. Salm, P. A., Scanziani, M., Malenka, R. C. & Nicoll, R. A. Distinct short-term plasticity at two excitatory synapses in the hippocampus. Proc. Natl Acad. Sci. USA 93, 13304–13309 (1996).
5. Nanou, E. & Catterall, W. A. Calcium channels, synaptic plasticity, and neuropsychiatric disease. Neuron 98, 466–481 (2018).
6. Grant, S. G. N. Synapse diversity and synapome architecture in human genetic disorders. Hum. Mol. Genet. 28, R219–R225 (2019).
7. Moradi, K. & Ascoli, G. A. Systematic data mining of hippocampal synaptic properties. In: Hippocampal Microcircuits A Computational Modeler’s Resource Book (eds Cutsuridis, V., Graham, B. P., Cobb, S., Vida, I.), 2 edn. (Springer International Publishing, 2019).
8. Ascoli, G. A. & Wheeler, D. W. In search of a periodic table of the neurons: axonal-dendritic circuitry as the organizing principle. Patterns of axons and dendrites within distinct anatomical parcels provide the blueprint for circuit-based neuronal classification. Bioessays 38, 969–976 (2016).
9. Rees, C. L., Moradi, K. & Ascoli, G. A. Weighing the evidence in Peters’ rule: does neuronal morphology predict connectivity? Trends Neurosci. 40, 63–71 (2017).
10. Wheeler D. W. et al. Hippocampome.org: a knowledge base of neuron types in the rodent hippocampus. eLife 4, (2015).
11. Moradi, K. & Ascoli, G. A. A comprehensive knowledge base of synaptic electrophysiology in the rodent hippocampal formation. Hippocampus 30, 314–331 (2020).
12. Morrison, A., Diesmann, M. & Gerstner, W. Phenomenological models of synaptic plasticity based on spike timing. Biol. Cybern. 98, 459–478 (2008).
13. Tsoodyks, M., Pawelzik, K. & Markram, H. Neural networks with dynamic synapses. Neural Comput. 10, 821–835 (1998).
14. Ecker, A. et al. Data-driven integration of hippocampal CA1 synaptic physiology in silico. Hippocampus 30, 1129–1145 (2020).
15. Lazebnik, Y. Can a biologist fix a radar?—Or, what I learned while studying apoptosis. Cancer Cell 2, 179–182 (2002).
16. Heidari, M., Jones, J. H. & Uzuner, O. Deep contextualized word embedding for text-based online user profiling to detect social bots on twitter. In: 2020 International Conference on Data Mining Workshops (ICDMW). (IEEE, 2020).
17. Heidari, M. & James, H. Uzuner, O. An empirical study of machine learning algorithms for social media bot detection. In: 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRANICS) (IEEE, 2021).
18. Fan, J. & Chow, T. Deep learning based matrix completion. Neurocomputing 266, 540–549 (2017).
19. Stulp, F. & Sigaud, O. Many regression algorithms, one unified model: a review. Neuroviet. Neurol. 69, 60–79 (2015).
20. Richards, B. A. et al. A deep learning framework for neuroscience. Nat. Neurosci. 22, 1761–1770 (2019).
21. Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S. & Acharya, U. R. Deep learning for healthcare applications based on physiological signals: a review. Comput. Methods Prog. Biomed. 161, 1–13 (2018).
22. Van Hulse, J. & Khoshgoftaar, T. Knowledge discovery from imbalanced and noisy data. Data Knowl. Eng. 68, 1513–1542 (2009).
23. Venkadesh, S., Komendantov, A. O., Wheeler, D. W., Hamilton, D. J. & Ascoli, G. A. Simple models of quantitative firing phenomenotypes in hippocampal neurons: Comprehensive coverage of intrinsic diversity. PLoS Comput Biol. 4, e1007462 (2019).
24. Beaulieu-Larocche, L. & Harnett, M. T. Dendritic spines prevent synaptic voltage clamp. Neuron 97, 75–82.e7 (2018).
25. Sacco, F., Fiumelli, H., Benfenati, F., Cancedda, I. & Barberis, A. Intracellular chloride concentration influences the GABAA receptor subunit composition. Nat. Commun. 3, 738 (2012).
26. Hamilton, D. J., White, C. M., Rees, C. L., Wheeler, D. W. & Ascoli, G. A. Molecular fingerprinting of principal neurons in the rodent hippocampus: a neuroinformatics approach. J. Pharm. Biomed. Anal. 144, 269–278 (2017).
27. White, C. M., Rees, C. L., Wheeler, D. W., Hamilton, D. J. & Ascoli, G. A. Molecular expression profiles of morphologically defined hippocampal neuron types: Empirical evidence and relational inferences. Hippocampus 30, 472–487 (2020).
28. Rees, C. L., White, C. M. & Ascoli, G. A. Neurochemical markers in the mammalian brain: structure, roles in synaptic communication, and pharmacological relevance. Curr. Med. Chem. 24, 3077–3103 (2017).
29. Busquets-Garcia, A., Bains, J. & Marsicano, G. CB1 receptor signaling in the brain: extracting specificity from ubiquity. Neuropsychopharmacology 43, 4–20 (2018).
30. Cossart, R. et al. Interneurons targeting similar layers receive synaptic inputs with similar kinetics. Hippocampus 16, 408–420 (2006).
31. Dumitriu, D., Cossart, R., Huang, J. & Yuste, R. Correlation between axonal morphologies and synaptic input kinetics of interneurons from mouse visual cortex. Cereb. Cortex 17, 81–91 (2007).
32. Jiang, X. et al. Principles of connectivity among morphologically defined cell types in adult neocortex. Science 350, aac9462 (2015).
33. Teyser, C., Wheeler, D. W., Sutton, N. & Ascoli, G. A. Comprehensive estimates of potential synaptic connections in local circuits of the rodent hippocampal formation by axonal-dendritic overlap. J. Neurosci. 41, 1665–1683 (2021).
34. Zeng, W. & So, H. C. Outlier-robust matrix completion via lp-minimization. IEEE Trans. Signal Process. 66, 1125–1140 (2018).
35. Soares, C., Trotter, D., Longtin, A., Beije, J. C. & Naud, R. Parsing out the variability of transmission at central synapses using optical quantal analysis. Front. Synaptic Neurosci. 11, 22 (2019).
36. Rosbroich, J., Trotter, D., Beninger, J., Toth, K. & Naud, R. Linear-nonlinear cascades capture synaptic dynamics. PLoS Comput. Biol. 17, e1008013 (2021).
37. Toth, K., Suares, G., Lawrence, J. J., Philips-Tansey, E. & McBain, C. J. Differential mechanisms of transmission at three types of mossy fiber synapse. *J. Neurosci.* 20, 8279–8289 (2000).
38. Oberlander, J. G. & Woolley, C. S. 17β-estradiol acutely potentiates glutamatergic synaptic transmission in the hippocampus through distinct mechanisms in males and females. *J. Neurosci.* 36, 2677–2690 (2016).
39. Fester, L. & Rune, G. M. Sexual neurosteroids and synaptic plasticity in the hippocampus. *Brain Res.* 1621, 162–169 (2015).
40. Higura-Matas, A. et al. Sex-specific disturbances of the glutamate/GABA balance in the hippocampus of adult rats subjected to adolescent cannabinoid exposure. *Pharmacology* 62, 1975–1984 (2012).
41. Moradi, K., Kaka, G. & Gharibzadeh, S. The role of passive normalization, voltage-gated channels and synaptic scaling in site-independence of somatic Dendritic Amplification in CA1 pyramidal neurons. *Neurosci. Res.* 73, 8–16 (2012).
42. Fatima-Shad, K. & Barry, P. H. Anion permeation in GABA- and glycine-gated channels of mammalian cultured hippocampal neurons. *Proc. Biol. Sci.* 265, 69–75 (1993).
43. Velumian, A. A., Zhang, L., Pennefather, P. & Carlen, P. L. Reversible inhibition of KIC and GABAergic cells in the rat hippocampal pyramidal neurons. *PLoS. Arch.* 433, 343–350 (1997).
44. Komendtov, A. O. et al. Quantitative firing pattern phenotyping of hippocampal neuron types. *Sci. Rep.* 9, 17915 (2019).
45. Morales, M., Wang, S. D., Diaz-Ruiz, O. & Jho, D. H. Cannabinoid CB1 receptor and serotonin 3 receptor subunit A (5-HT3A) are co-expressed in GABA neurons in the dentate gyrus. *J. Comp. Neurol.* 468, 205–216 (2004).
46. Akima, H. A new method of interpolation and smooth curve fitting based on local procedures. *J. ACM* 17, 589–602 (1970).
47. Savanthrapadian, S. et al. Synaptic properties of SOM- and CCK-expressing inhibitory neurons. *Front. Neurosci.* 200, 69–74 (2020).
48. Prechelt, L. Early stopping—but when? In: *Chaos Solitons Fractals*, 109853 (2020).
49. Bartos, M., Vida, I., Frotscher, M., Geiger, J. R. & Jonas, P. Rapid signaling at hebbian LTP in CA1 stratum oriens interneurons: interactions between group I metabotropic glutamate receptors and M1 muscarinic receptors. *J. Neurosci.* 35, 13542–13554 (2015).
50. Mercer, A., Eastlake, K., Trigg, H. L. & Thomson, A. M. Local circuitry involving parvalbumin-positive basket cells in the CA2 region of the hippocampus. *Hippocampus* 22, 43–56 (2012).

Acknowledgements
We acknowledge the contributions of our colleagues Christopher Rees, Diek Wheelker, Nate Sutton, Jeffrey Kopsick, Sarojini Attili, Carolina Tecuatl, Kayvan Bijari, Masood Akram, Sumit Nanda, Ketan Mehta, Srivadi Polavararam, Siva Venkadesh, Alexander Komendantov, David Hamilton, Rubén Armañanzas, and Patricia Maraver who kindly shared their valuable feedback at various stages of the project; high school or undergraduate students Sung Joon Park, Alisha Compton, Saisruthi Kannan, Ridha Rahim, Samantha Barta, Payal Panchal, Deepika Rao, Jacinta Das, Maham Saleem, Manuel Carrasco, and Zaid Alzamani who contributed to data mining and development of tools; Engage Digitizer software developer Mark Mitchell who patiently finalized the manuscript. G.A.A. supervised every aspect of the study and was responsible for grant funding acquisition and management. We acknowledge the contributions of our colleagues Christopher Rees, Diek Wheelker, Nate Sutton, Jeffrey Kopsick, Sarojini Attili, Carolina Tecuatl, Kayvan Bijari, Masood Akram, Sumit Nanda, Ketan Mehta, Srivadi Polavararam, Siva Venkadesh, Alexander Komendantov, David Hamilton, Rubén Armañanzas, and Patricia Maraver who kindly shared their valuable feedback at various stages of the project; high school or undergraduate students Sung Joon Park, Alisha Compton, Saisruthi Kannan, Ridha Rahim, Samantha Barta, Payal Panchal, Deepika Rao, Jacinta Das, Maham Saleem, Manuel Carrasco, and Zaid Alzamani who contributed to data mining and development of tools; Engage Digitizer software developer Mark Mitchell who patiently fixed the bugs and added requested features. This study was supported in part by National Institute of Health (NIH) grants R01NS39600 and U01MH114829.

Author contributions
K.M. and G.A.A. designed the study. K.M., G.P.M., and Z.A. mined the data. K.M. and Z.A. digitized and simulated synaptic signals. K.M. and G.A.A. conducted the machine learning and statistical analysis; K.M., G.P.M., M.L., and G.A.A. wrote the initial draft. K.M and Z.A. prepared the figures. M.L. and G.A.A. made the data set. K.M. and Z.A. finalized the manuscript. G.A.A. supervised every aspect of the study and was responsible for grant funding acquisition and management.

Competing interests
The authors declare no competing interests.

Ethical approval
No new human or animal experimental data were collected for the current study. The animal dataset comes from a curated database that was described in Moradi and Ascoli. That publication does not include any ethics information because it was itself derived by a literature meta-analysis and not a primary study. The mentioned meta-analysis involves 160 articles (Supplementary Data 2), each with their own individual ethics information.
