A sequential two-step priming scheme reproduces diversity in synaptic strength and short-term plasticity

Kun-Han Lin, Holger Taschenberger, and Erwin Neher

Contributed by Erwin Neher; received May 9, 2022; accepted July 13, 2022; reviewed by Zoltan Nusser and Samuel Young Jr.

Glutamatergic synapses display variable strength and diverse short-term plasticity (STP), even for a given type of connection. Using nonnegative tensor factorization and conventional state modeling, we demonstrate that a kinetic scheme consisting of two sequential and reversible steps of release–machinery assembly and a final step of synaptic vesicle (SV) fusion reproduces STP and its diversity among synapses. Analyzing transmission at the calyx of Held synapses reveals that differences in synaptic strength and STP are not primarily caused by variable fusion probability ($p_{\text{fusion}}$) but are determined by the fraction of docked synaptic vesicles equipped with a mature release machinery. Our simulations show that traditional quantal analysis methods do not necessarily report $p_{\text{fusion}}$ of SVs with a mature release machinery but reflect both $p_{\text{fusion}}$ and the distribution between mature and immature priming states at rest. Thus, the approach holds promise for a better mechanistic dissection of the roles of presynaptic proteins in the sequence of SV docking, two-step priming, and fusion. It suggests a mechanism for activity-induced redistribution of synaptic efficacy.

synaptic transmission | short-term plasticity | synaptic vesicle priming | calyx of Held | numerical simulation

Chemical synapses change their strength during repetitive use in a synapse type-specific and activity-dependent manner. Such modifications occur on several timescales and define dynamic properties of synaptic networks (1, 2). Elucidating the biophysical mechanisms of synaptic plasticity is essential to understand information processing in circuits (3, 4). Kinetic schemes of synaptic transmission and plasticity provide a theoretical framework to mechanistically and quantitatively interpret functional mechanisms of synaptic plasticity is essential to understand information processing in circuits (3, 4). Kinetic schemes of synaptic transmission and plasticity provide a theoretical framework to mechanistically and quantitatively interpret functional mechanisms of synaptic plasticity.

Short-term changes of synaptic strength such as paired-pulse facilitation (PPF) and short-term depression (STD) have been ascribed to changes in fusion probability ($p_{\text{fusion}}$) of synaptic vesicles (SVs) and/or changes in occupancy of presynaptic release sites (6). Some basic features of short-term plasticity (STP) are captured by a simple scheme posulating one kind of release site to which SVs are recruited, possibly in a Ca$^{2+}$-dependent manner, before being able to fuse upon action potential (AP) arrival (7, 8). However, numerous observations, including multiple kinetic components of STD (9) and its recovery (10, 11) and diverse STP even among synapses of a given type (12–15) are not easily accounted for by such a simple model.

To more faithfully reproduce the multifaceted features of STP, different multiple-state and/or multiple-site schemes of transmitter release have been proposed (16–24), which include parallel schemes in which more than one kind of SV can bind to one or more kinds of release sites, and sequential schemes in which SVs migrate between different kinds of release sites or states of maturation.

Motivated by converging evidence from molecular biology (25–29), electrophysiology (30–33), live-cell imaging (34, 35), and electron microscopy (EM) (36, 37) emphasizing the reversibility and multistep nature of the priming process, we explore here whether a recently proposed single-site multiple-state scheme of priming and fusion (38) can reproduce variable synaptic strength and diverse STP observed at the calyx of Held, a mammalian glutamatergic model synapse. The proposed kinetic scheme in its basic form (Fig. 1A and B) assumes that SVs reversibly dock to a single type of release site and undergo two sequential priming steps to become fusion competent. Considering ultrastructural evidence for distinct docking states (36, 39–42), we refer to the two states and to the SVs residing in those states as loosely (LS) or tightly (TS) docked and SV$_{\text{LS}}$ or SV$_{\text{TS}}$, respectively (see SI Appendix, Table S1 for a list of abbreviations).

By analyzing AP-evoked EPSC (eEPSC) trains elicited by a wide range of presynaptic firing frequencies with a combination of nonnegative tensor factorization (NTF)

Significance

Central nervous system synapses are diverse in strength and plasticity. Short-term plasticity has traditionally been evaluated with models postulating a single pool of functionally homogeneous fusion-competent synaptic vesicles. Many observations are not easily explainable by such simple models. We established and experimentally validated a scheme of synaptic vesicle priming consisting of two sequential and reversible steps of release–machinery assembly. This sequential two-step priming scheme faithfully reproduced plasticity at a glutamatergic model synapse. The proposed priming and fusion scheme was consistent with the measured mean responses and with the experimentally observed heterogeneity between synapses. Vesicle fusion probability was found to be relatively uniform among synapses, while the priming equilibrium at rest of mature versus immature vesicle priming states differed greatly.

Author contributions: K.-H.L., H.T., and E.N. designed research; K.-H.L. performed research; K.-H.L., H.T., and E.N. analyzed data; and H.T. and E.N. wrote the paper.

Reviewers: Z.N., Kiserlei Orvostudomanyi Kutatointezet; and S.Y., The University of Iowa.

The authors declare no competing interest.

Copyright © 2022 the Author(s). Published by PNAS.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

1K.-H.L. and H.T. contributed equally to this work.

To whom correspondence may be addressed. Email: eneher@mpinat.mpg.de or taschenberger@mpinat.mpg.de.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2207987119/-/DCSupplemental.

Published August 15, 2022.
and conventional state modeling, we reached a number of conclusions that provide views on the mechanisms of neurotransmitter release and STP. First, approximately 80% of available release sites are occupied at rest by primed SVs, which can be either in the LS or the TS. Second, different initial strength and diverse STP among synapses is primarily due to the variable relative abundance of SV_{TS} over SV_{LS}, while p_{fusion} is quite uniform with a high value of ∼0.4. Third, for frequencies of 5 to 20 Hz, steady-state release rates scale roughly linearly with presynaptic firing rates, thus maintaining largely frequency-invariant synaptic strength. Finally, at frequencies ≥50 Hz, additional kinetic features control release, such as an increase in p_{fusion} during trains, a speed-up of the priming process, and a decline of SV subpool occupancies reducing steady-state release.

Our numerical simulations, which mostly use experimentally determined or NTF-constrained model parameters, faithfully reproduce STP at calyx synapses over a wide range of activity levels and therefore provide a valuable framework for the mechanistic and quantitative interpretation of experimentally induced STP alterations. By emphasizing the multistep nature and reversibility of the priming process, the sequential two-step priming scheme suggests a mechanism for creating functional diversity among synapses and for an activity-induced redistribution of synaptic efficacy during AP trains.

**Results**

To validate the two-step priming scheme, we chose the following three-step approach. First, we acquired eEPSCs evoked by regular stimulus trains (0.5–200 Hz) from an ensemble of 35 rat calyx synapses. In addition, 100 and 200 Hz eEPSC trains preceded by two or four stimuli at 10 Hz were recorded. Second, all eEPSC train peaks were converted to quantal content, and such data from five to 200 Hz trains were subsequently subjected to NTF analysis. Third, model parameters were then initialized with values derived from NTF analysis or from analytical expressions regarding model predictions for, e.g., paired-pulse-ratio (PPR) and steady-state release at low-frequency stimulation (Eqs. 22–24 [equations referred to here and in the following are provided in the SI Appendix]). Subsequently, model fits were optimized by trial-and-error parameter variation to closely reproduce both the respective average eEPSC train response for each stimulation frequency (f_{stim}) and the NTF-derived base functions (BF). Including BFs in this optimization procedure ensures consistency of the model not only with mean train responses but also with the heterogeneity among synapses.

Estimating release from peak amplitudes neglects asynchronous release, which builds up during interstimulus intervals (ISIs) and decays after stimulation (43). However, this release component is small and decreases further during calyx maturation (44). We focused on presynaptic mechanisms regulating STP by 1) choosing P14–16 calyx synapses, which are little affected by AMPAR saturation and desensitization (45, 46), and 2) recording eEPSCs in the presence of 1 mM kynurenic acid (kyn) to alleviate remaining postsynaptic effects (45). Presynaptic inhibition of release via metabotropic GluRs is reportedly low at this age (47), which leaves the modulation of p_{fusion} and the dynamic regulation of priming as major determinants of STP.

**Experimental eEPSC train data: Mean time courses and variability among calyx synapses.** Fig. 2 A and B illustrate eEPSCs and the time courses of mean quantal content (m) in response to stimulus trains consisting of 15 APs (0.5, 1, and 2 Hz) or 40 APs (5–200 Hz). For a given synapse, m estimates for the initial eEPSCs (m_{1}) were similar across all f_{stim}. However, the mean m_{1} across all f_{stim} varied nearly 10-fold between synapses (77–739 SVs, coefficient of variation (CV) = 0.42; Fig. 2D). The average over all 35 mean m_{1} values was 377 ± 27 SV (Fig. 2B). During trains, the average m_{1} decreased monotonically toward a depressed steady state for all but the highest f_{stim}. During 200 Hz stimulation, net facilitation was observed for the average response—i.e., the PPR (m_{2}/m_{1}) was on average >1 (SI Appendix, Fig. S1A). PPR, when plotted against ISI, converts from facilitation into depression with a time constant of ∼21 ms (SI Appendix, Fig. S1A) (48). Plotting PPR_{200 Hz} as a function of m_{1} for individual synapses revealed large diversity and a negative correlation—i.e., synapses with larger m_{1} predominantly showed paired-pulse depression while those with smaller m_{1} often exhibited PPF (Fig. 2D). Such correlation is usually interpreted as an indication of heterogeneous initial p_{fusion} (p_{fusion1}), because strong depression in synapses with high p_{fusion1} would occlude PPF (49). The two-step priming scheme (Fig. 1 A and B) allows an alternative view, as detailed in SI Appendix. It provides an approximate expression for calculating p_{fusion1} from PPR and STD during 10 Hz trains (Eq. 31) for each synapse. Plotting such p_{fusion1} estimates versus the respective m_{1} values revealed only weak correlation (Fig. 2E). The mean p_{fusion1} amounted to 0.43 with a CV of 0.20. The latter was substantially smaller than the CV of m_{1} values, indicating that the variability of p_{fusion1} is not the principal cause for heterogeneous synaptic
strength. It is rather due to a variable size of the subpool of fusion-competent SVs at rest (SP\textsubscript{TS,rest}), which is readily calculated for each synapse as the ratio \(m\text{\textsubscript{TF}}/m\text{\textsubscript{fusion}}\). Plotting \(SP\text{\textsubscript{TS,rest}}\) estimates as a function of the respective \(m\text{\textsubscript{1}}\) values revealed a strong linear correlation with an expected slope of \(\sim 1/0.43\) (Fig. 2F). The estimated mean \(SP\text{\textsubscript{TS,rest}}\) (880 ± 57 SVs) had a CV of 0.38, which was close to the CV of \(m\text{\textsubscript{1}}\) values.

In sum, a large heterogeneity among synapses with respect to their \(SP\text{\textsubscript{TS}}\) size at rest can explain strong variability in \(m\text{\textsubscript{1}}\), while \(p\text{\textsubscript{fusion}}\) varies only little.

**NTF analysis provides \(p\text{\textsubscript{fusion}}\) estimates and yields constraints for SV subpool sizes and priming kinetics.** We next derived suitable initial guesses for model parameters from NTF analysis. NTF decomposes complex data sets into a linear combination of components (50) and was recently adapted for the analysis of eEPSC trains (51). The algorithm considers time courses of quantal contents during trains as superpositions of contributions by two or more types of signal sources and assumes that their differential relative contributions account for STP diversity among synapses. In the context of the model (Fig. 1A), these sources correspond to the release contributions by SVs residing in certain states (such as TS and LS) prior to stimulation. NTF analysis returns the time courses of individual contributions as BFs (51). For a given \(f\text{\textsubscript{ts zad}}\) BFs are the same for all synapses. They are normalized to a cumulative sum of 1, such that the product of a BF and the corresponding synapse-specific train quantal content (\(M\)) represents the time-resolved quantal release of that component for a given synapse (Fig. 3B and SI Appendix, Fig. S2 A and B). Two-component NTF fits provide BFs for release contributed by preexisting \(SV\text{\textsubscript{TS}}\) (BF\textsubscript{TS}; Fig. 3A) and for the combined remaining release originating from SVs, which had been either loosely docked at stimulation onset or were newly recruited during the train (BF\textsubscript{LS,RS}; Fig. 3B). The latter release component can be decomposed by subsequent 3-component NTF analysis (Fig. 3C and SI Appendix, Fig. S2 A and B). In short, NTF provides time courses of release contributions of SVs or sites, which had been in one of the states of the model (ES, TS, and LS; Fig. 1A) at stimulation onset.

NTF analysis cannot provide initial guesses for all model parameters, but it is very instrumental in constraining some of them. Two-component NTF, which is robust (51), separates release contributed by preexisting \(SV\text{\textsubscript{TS}}\) from other contributions. BF\textsubscript{TS} decay rapidly and approach zero after \(~5\) APs (Fig. 3 A). Provided that all preexisting \(SV\text{\textsubscript{TS}}\) are consumed, their initial value represents \(p\text{\textsubscript{fusion}}\) (0.39 ± 0.004 when averaged over all six \(f\text{\textsubscript{ts zad}}\)), and the train quantal content \(MT\text{\textsubscript{TS}}\) associated with BF\textsubscript{TS} represents an estimate for the subpool \(SP\text{\textsubscript{TS}}\) at

---

**Fig. 2.** STP in response to 0.5–200 Hz stimulus trains in post hearing-onset calyx of Held synapses. (A) Sample eEPSCs obtained from a strongly depressing (top) and a facilitating (bottom) synapse in response to 200 Hz (Left), 20 Hz (Middle), and 2 Hz (Right) stimulation. Only the initial 15 eEPSC trains are superimposed for the 2 Hz and 20 Hz eEPSC trains. Each trace represents an average of three repetitions. (B) Mean quantal content (\(m\text{\textsubscript{1}}\)) plotted against stimulus index \(i\) (\(B, i\)) or time (\(B, ii\)) for each eEPSC. Train consists of only 15 stimuli for the lowest three frequencies. The timing of eEPSC was offset by one ISI in (\(B, i\)) for clarity. Note logarithmic time axis in (\(B, ii\)). (C) Estimating the fast releasing pool (FRP) of SVs from eEPSC trains evoked by 50, 100, and 200 Hz stimulation which provided three FRP estimates that are uncorrected for incomplete pool depletion. The relationship between the three 1/FRP values and their respective ISIs was subsequently extrapolated to infinite \(f\text{\textsubscript{ts zad}} = 0\) ms to obtain a mean FRP value that is corrected for incomplete pool depletion (Inset). (D) PPRs \((m\text{\textsubscript{1}}/m\text{\textsubscript{1}})\) for 200 Hz eEPSC trains negatively correlate with initial quantal content \(m\text{\textsubscript{1}}\), which varies approximately 10-fold among calyx synapses (73–728 SVs). The gray shaded region indicates PPR >1. (E and F) Predictions for \(p\text{\textsubscript{fusion}}\) (E) and \(SP\text{\textsubscript{TS,rest}}\) (F) for individual synapses obtained from their respective 10 Hz PPR and \(D\text{\textsubscript{ts zad}}\) values according to SI Appendix, Eq. 31.
rest \((SP_{TS,rest}; \text{average } 961 \pm 68 \text{ SVs})\). BF\(_{TS}\)'s for \(f_{\text{stim}}\) of 5 to 20 Hz are strikingly similar when plotted against stimulus number (Fig. 3 A, i). They decay exponentially, indicating nearly constant \(p_{\text{fusion}}\) throughout trains at these frequencies. BF\(_{TS}\)'s for \(\geq 50\) Hz deviate from this pattern (Fig. 3 A, ii). Their second values are larger, followed by a steeper decline, which is especially prominent at \(50\) Hz. Quantitative analysis of the time courses of \(p_{\text{fusion}}\) during trains as derived from BF\(_{TS}\) (51) indicates a small decrease in \(p_{\text{fusion}}\) for \(f_{\text{stim}}\) of 5 to 20 Hz, while for \(\geq 50\) Hz \(p_{\text{fusion}}\) increases, consistent with the observed net facilitation (Fig. 3D). Fig. 3B shows the time courses for quantal release contributed by those SVs, which were not in the TS state at stimulus onset (Release\(_{TS,RS}\)). Again, for 5 to 20 Hz, these time courses are strikingly similar when plotted against the stimulus number, indicating similar contributions to release regardless of ISI duration (Fig. 3 B, i).

Three-component NTF decomposition allows the separation of release contributions from preexisting \(SV_{LS}\)s (Fig. 3 C) and newly recruited SVs (SI Appendix, Fig. S2A) (51). Provided that all preexisting \(SV_{LS}\)s are consumed, the train quantal content (\(M_{TS,\delta}\) associated with the BF\(_{LS}\) represents an estimate for \(SP_{TS}\) at rest \((SP_{LS,rest}; \text{average } 1,078 \pm 75 \text{ SVs})\). BF\(_{LS}\) start at a value very close to zero since preexisting \(SV_{LS}\)s cannot fuse during the first AP but first need to undergo the LS \(\rightarrow\) TS transition. For all frequencies, BF\(_{LS}\) quickly increase to a maximum value after 2 to 3 APs before decaying exponentially to near zero. Again, the time courses of BF\(_{LS}\) for 5 to 20 Hz are nearly indistinguishable (Fig. 3 C, ii). As pointed out previously (51) and formally proven here (Eqs. 10–20), this finding is consistent with a transition of a constant fraction \((s_2)\) of \(SP_{TS}\) to \(SP_{TS}\) subsequent to an AP. The average BF\(_{LS}\) for 5 to 20 Hz trains can be used to obtain an estimate for \(s_2\) (0.11; Fig. 3 C, iii).

At steady state, the loss from \(SP_{TS}\) due to SV fusion is compensated by an equal number of SVs replenishing \(SP_{TS}\). Because the time course of BF\(_{LS}\) (Fig. 3 C, i) along with that of the release contributed by newly recruited SVs (Release\(_{RS}\); SI Appendix, Fig. S2 A, i) are frequency invariant for \(f_{\text{stim}} = 5\) to 20 Hz, subpool occupancies approach very similar steady-state values regardless of ISI durations. That way, SVs are supplied to \(SP_{TS}\) at a rate linearly increasing with \(f_{\text{stim}}\) which yields a frequency invariant steady-state quantal content \((m_s)\) at 5 to 20 Hz (SI Appendix, Fig. S1 B–D) (52, 53). We refer to this feature as balanced priming.

![Figure 3](https://www.pnas.org/content/10.1073/pnas.2207987119)
Simulations using the basic model confirm predictions derived from NTF analysis. After initializing model parameters (Fig. 1B) as described, numerical simulations reproduce steady-state depression \( (D_m = m_j/m_i) \) for \( f_{stim} = 1 \) to 20 Hz quite well (Fig. 4D, dotted trace). We assumed for both priming rate constants \( k_1 \) and \( k_2 \) fixed resting values \( (k_{1rest}, k_{2rest}) \) and linear slopes \( \sigma_1 \) and \( \sigma_2 \) to describe their \( Ca^{2+} \) dependence (Eqs. 4, 5). The resting size of subpools, \( SP_{TSrest} \) and \( SP_{SSrest} \), are thus determined by \( k_{1rest}, \theta_1, k_{2rest} \), and \( \beta_2 \) (Eqs. 1–5). During stimulation, \( SP_{TS} \) partially depletes, such that release decreases toward a steady-state value \( m_s \) for which SV consumption and replenishment are at equilibrium (Fig. 4D). At \( f_{stim} \leq 2 \) Hz, this balance depends on all the parameters listed here, and increasing \( f_{stim} \) results in reduced replenishment and lower \( m_s \) due to shorter ISIs (Fig. 4D). However, at \( f_{stim} \) of 5 to 20 Hz, the \( Ca^{2+} \)-dependent increments of \( k_1 \) and \( k_2 \) above their resting values dominate the priming rates. Unpriming, determined by \( b_1 \) and \( b_2 \), becomes negligible, and the net movement of SVs along the kinetic scheme occurs nearly exclusively in the priming direction. The resulting steady-state occupancy of subpools is dominated by \( pfusion \) and the integrals.
s1 and s2 over priming rates between consecutive APs (Eqs. 10, 12, and 24), the latter representing fractions of upstream subpools converted per AP to the corresponding downstream subpools. s1 and s2 are related to the slopes σ1 and σ2 according to Eq. 12. They are independent of f fuse as long as [Ca2+] transients are short relative to ISIs and [Ca2+]−dependent priming dominates. Thus, mT tends toward a nearly frequency-independent plateau value (Eq. 24) since both SV consumption and replenishment increase approximately linearly with frequency (Fig. 4 D, dotted trace) (52, 53). Recovery from STD following 10 Hz stimulation is well approximated by single exponentials with time constants ~4 s (54), which can be reproduced in simulations by appropriate selection of kₜₗₑ₉ and kₜₑ₉ₑ₉.

In sum, STP during 1 to 20 Hz eEPSC trains is well described by a balanced priming scheme, as suggested by NTF, in which each AP transfers constant fractions (s1 and s2) of SVs from the respective upstream subpools to SPₜₛ and SPₜₛ and releases an almost constant fraction of SPₜₛ (= p fuse). This leads to increasing STD (decrease of mₖ) for increasing f fuse up to 5 Hz, when Ca2+-independent rates and Ca2+-dependent rates are of similar magnitude. However, for f fuse ≥ 5 Hz, when Ca2+-dependent rates dominate, occupancy of both SPₜₛ and SPₜₛ is nearly frequency-independent, leading to relatively stable mₖ despite increasing f fuse. For f fuse > 20 Hz, however, this simple scheme will have to be extended, as detailed below.

Multiple mechanisms of release facilitation at stimulus frequencies ≥50 Hz. For f fuse ≥ 50 Hz, large deviations are evident between experimental data and the basic model, which therefore needs to be extended to reproduce STP for f fuse up to 200 Hz. Inspection of the respective BFs reveals several changes at high f fuse: 1) p fuse increases during trains, which is reflected in elevated second and third values of BFₜₛ, followed by a more rapid decay (Fig. 3 A, ii and D); 2) release contributed by those SVs not residing in TS prior to stimulation develops a peak around the third to fifth AP before decaying to steady-state levels lower than those at ≤20 Hz (Fig. 3 B, ii and C, ii); and 3) steady-state release contributed by newly recruited SVs decreases with increasing f fuse causing stronger steady-state depression (Figs. 3 B, ii and 4D and SI Appendix, Fig. S2 A, ii). These changes in synapse behavior, which are likely caused by a summation of [Ca2+] transients or an incomplete relaxation of inter- motion during high activity levels.

Saturation of priming causes increased steady-state depression during high activity levels. At f fuse ≥50 Hz, a progressive decline of release after the second or third stimulus toward a lower mₖ is observed (Figs. 4 A, ii and 5 B). It results in a saturation of the relationship between release rate and f fuse (SI Appendix, Fig. S1C). We explored two options for modeling this: 1) a saturation of the priming rate with increasing f fuse and 2) a delayed availability of recently used sites for SV docking by introduction of a refractory release site state (58, 59).

As the simplest case, we describe here the replacement of the linear relationship between k₂ and [Ca2+] (Eq. 4) by a Michaelis– Menten (MM) type saturating relationship (Eq. 42). This introduces a parameter Kₐₛ, which is the [Ca2+] at half-maximum k₂. A Kₛₐ of 280 nM and a slight adjustment of other parameters adequately reproduces experimental data for all f fuse (Fig. 4 A and SI Appendix, Table S2). Note that the simulated mean steady-state depression Dₘₑₙ = mₖ/mₖₑ₉ exactly superimposes on experimental data over two orders of magnitude of f fuse (Fig. 4 D).

Having established an adequate fit to the experimental data, we then examined time courses of model quantities such as effective [Ca2+]ₜₛ, p fuse, and subpool occupancies (Figs. 4 B and C and 3D). At 10 Hz stimulation, the occupancy of SPₜₛ remains relatively constant during the train, while SPₜₛ rapidly depletes to a level of 0.310 times its resting value. This decay is the major cause of observed STD at 10 Hz (mₖ/mₑ₉ = 0.312 ± 0.013). At 200 Hz stimulation, both SPₜₛ and SPₜₛ strongly deplete, and the depletion of the former is a consequence of the MM type saturation of k₂. Depletion of subpools is partially compensated by an increase in p fuse (Fig. 3 D). Model predictions for steady-state subpool sizes and depression are plotted against f fuse in Fig. 4 D.

We assume that each AP converts a certain fraction k of SPₜₛ into such labile SVₜₛₚ₞, which constitute SPₜₛ (Eqs. 43–46). They have the same p fuse as SVₜₛₚ₞, but their backward transition LS ← TSL (b₃) is more rapid than that of LS ← TS (b₅). Thus, they contribute “extra” release to all responses of a high-frequency train except eEPSC, thereby enhancing PPR. They do not contribute to release, when ISIs are much longer than their mean decay time (1/b₅). A labile priming state can be implemented in series between LS and TS or as a parallel branch to the LS → TS transition. For both options, model parameters can be found, which reproduce STP accurately, even when the p fuse of SVₜₛₚₚₚₙ is that of SVₜₛₚₚ is constrained to have the same value. For brevity, we describe here only the version with the parallel branch (Fig. 1 C). SPₜₛ and SPₜₛ are incremented and decremented, respectively, following each AP by an amount of k · SPₜₛ. During ISIs, the time courses of subpool occupancies including SPₜₛ are described by the differential equations Eqs. 43–46. By introducing this extension of a TSL together with the facilitation of p fuse, we were able to numerically simulate release during high-frequency trains with high fidelity (Fig. 5 B) and identified k = 16% and b₅ = 11.1 s⁻¹ as best parameters (SI Appendix, Table S2). These numbers are compatible with recent EM data showing that the number of SVs tightly docked at the AZ is transiently increased 5 ms after an AP but has relaxed back to normal when flash-freezing is initiated 100 ms later (37). Similar shifts between loosely and tightly docked SVs were reported to be associated with the induction of long-term potentiation (LTP) in hippocampal synapses (39) as well as with the beta-adrenergic modulation of parallel fiber LTP in the cerebellum (40).

In sum, release facilitation during the onset of high-frequency trains is generated by an increase in p fuse by a more rapidly replenished SPₜₛ due to accelerated priming when global [Ca2+] summates, and further by a transient filling-up of SPₜₛ.
This generates either pronounced net facilitation (Fig. 5A and B) and relative steady-state depression (Fig. 5C and D). Simulated BFs are shown in Fig. 5E and F. Contributions of simulated NTF components to total release at 200 Hz trains were simulated using standard values for all model parameters except for $b_{2}$, which was either increased ($C$, i and $D$, i) or decreased ($C$, ii and $D$, ii) such that the fraction $SPTS_{TS,rest}/(SPTS_{TS,rest} + SPTS_{rest})$ was reduced to $\sim 20\%$ or enhanced to $\sim 74\%$, respectively. The red dotted trace in $C$, ii represents the $m_{j}$ time course for the simulated 200 Hz train shown in $C$, i. (D) Simulated contributions to release during 200 Hz stimulation by preexisting $SV_{TS}$ (solid black) and by preexisting $SV_{TS}$ (dotted black) or newly recruited SVs (dashed black). Simulated total release ($m_{j}$, solid red) is shown for comparison. (E) Ratios $m_{j}\gamma_{m}$ (PPR) and $m_{j}\gamma_{m}$ ($E$, i) and relative steady-state depression ($m_{j}\gamma_{m}$, $E$, ii) plotted versus the relative fraction of $SPTS_{TS,rest}$ for 14 simulations similar to those shown in (C) and (D). Either unpriming rate constant $b_{2}$ or priming rate constant $b_{1}$ for the LS $\rightarrow$ TS transition were increased or decreased to generate relative $SPTS_{TS}$ fractions in the range from 0.2 to 0.5. Note that $\gamma_{m}SPTS_{TS,rest}$ was kept at 0.39 and standard values were used for all other model parameters. The gray shaded region in (E, i) indicates ratios $>1$.

As an alternative to a saturation of $k_{1}$, we introduced an empty but refractory release site state (ERS) ($59$–$61$). In this case, release sites vacated following SV fusion are not instantaneously converted into empty sites available for SV docking but shift into ERS, from which they become available for SV docking with a rate constant $b_{3}$ (Fig. 1C, i). A $b_{3}$ of $3.6\ s^{-1}$ and a slight optimization of other parameters (SI Appendix, Table S2) resulted in model predictions, which for our standard parameter set were hardly discernible from those of the alternative approach of using an MM–type saturation of $k_{1}$ (Fig. 1C, ii). Significant differences between the two types of models were only observed when the occupancy of release sites at rest was reduced below 70%.

In sum, decreasing steady-state release at $f_{stim} \geq 50 Hz$ indicates a saturation of the pool replenishment process, which can be simulated equally well by a saturation of the Ca$^{2+}$-dependent acceleration of $k_{1}$ or else by introducing a refractory state of release sites, which after their use become available for new SV docking with first-order kinetics.
STP induced by conditioning stimulation and accelerated eEPSC recovery after depleting high-frequency trains. We next studied STP during 100 and 200 Hz eEPSC trains preconditioned by two or four APs delivered at 10 Hz, which causes pronounced release facilitation (62). Two examples for eEPSCs elicited by such a stimulus pattern, which are part of our standard NTF protocol, are shown in Fig. 6 A and B. Nearly all synapses showed more or less pronounced facilitation following 10 Hz preconditioning (Fig. 6 C). The conditioning low-frequency stimulation depletes SPTS thereby exposing release facilitation at the onset of the high-frequency train due to the increase in $P_{\text{fusion}}$ and because of the rapid conversion of SVTLS to fusion-competent SVTSLS and SVTSLSS. The agreement between model predictions and average data was remarkable (Fig. 6 D). The conditioning release facilitation is sensitive to the $[\text{Ca}^{2+}]$ buildup during trains, which is determined by the decay of individual $[\text{Ca}^{2+}]$ transients. At $\leq 20$ Hz, when individual $[\text{Ca}^{2+}]$ transients do not overlap, the $[\text{Ca}^{2+}]$-dependence of $k_1$ and $k_2$ is relevant only in terms of the integral over transients—i.e., large transients with short duration are as effective in promoting priming as small transients with correspondingly longer duration. At $\geq 50$ Hz, however, individual $[\text{Ca}^{2+}]$ transients summate. During stimulus trains, the increased rate of priming is balanced by release. After stimulation, refilling of partially depleted subpools continues at an increased rate until $[\text{Ca}^{2+}]$ has decayed back to $[\text{Ca}^{2+}]_\text{rest}$. Lengthening the decay of the $[\text{Ca}^{2+}]$ transient at the expense of its amplitude shifts a larger proportion of its priming-promoting effect into the early recovery time course. In simulations, the amount of accelerated eEPSC recovery can be adjusted to match experimental data by varying the value of the model parameter $\tau_{\text{Ca}}$ (SI Appendix, Table S2). An example for a processes overlap: a fast drop of $P_{\text{fusion}}$ from a facilitated value back to $P_{\text{fusion,1}}$, a rapid decrease of $SP_{\text{TS}}$, and an increase in $SP_{\text{TS}}$ due to accelerated priming while $[\text{Ca}^{2+}]$ is elevated. The net result may either be an accelerated recovery or a transient decrease of synaptic strength, depending on the relative magnitudes of these processes and their kinetics (63). $[\text{Ca}^{2+}]$-accelerated priming is sensitive to the $[\text{Ca}^{2+}]$ buildup during trains, which is determined by the decay of individual $[\text{Ca}^{2+}]$ transients. At $\leq 20$ Hz, when individual $[\text{Ca}^{2+}]$ transients do not overlap, the $[\text{Ca}^{2+}]$-dependence of $k_1$ and $k_2$ is relevant only in terms of the integral over transients—i.e., large transients with short duration are as effective in promoting priming as small transients with correspondingly longer duration. At $\geq 50$ Hz, however, individual $[\text{Ca}^{2+}]$ transients summate. During stimulus trains, the increased rate of priming is balanced by release. After stimulation, refilling of partially depleted subpools continues at an increased rate until $[\text{Ca}^{2+}]$ has decayed back to $[\text{Ca}^{2+}]_\text{rest}$. Lengthening the decay of the $[\text{Ca}^{2+}]$ transient at the expense of its amplitude shifts a larger proportion of its priming-promoting effect into the early recovery time course. In simulations, the amount of accelerated eEPSC recovery can be adjusted to match experimental data by varying the value of the model parameter $\tau_{\text{Ca}}$ (SI Appendix, Table S2). An example for a
simulation of a stimulation pattern probing recovery is given in SI Appendix, Fig. S4. It should be noted, though, that experimentally observed decay time courses of global \([\text{Ca}^{2+}]\) are often biphasic (63–65). More detailed simulation of the \(\text{Ca}^{2+}\) signal relevant for controlling \(\text{Ca}^{2+}\)-dependent priming (i.e., effective \([\text{Ca}^{2+}]\)) may improve the accuracy of model predictions.

**Discussion**

We analyzed eEPSC trains elicited by a wide range of presynaptic firing frequencies using a combined electrophysiological and modeling approach that was aided by nonnegative tensor factorization (51). Our analysis leads to four principal conclusions: 1) The experimental data are well compatible with a reversible priming process leaving \(~10\%\) of the release sites empty at rest while the remaining \(80\%\) are occupied by SVs, which are in one of two states (LS and TS, constituting subpools \(\text{SP}_{\text{LS}}\) and \(\text{SP}_{\text{TS}}\) respectively). Both subpools equilibrate dynamically during presynaptic activity. 2) Different initial strength and diverse STP among calyx synapses is primarily due to variable \(\text{SP}_{\text{LS}}\) and \(\text{SP}_{\text{TS}}\) sizes at rest. Functional diversity across all synapses is consistent with relatively uniform \(P_{\text{fusion}}\) despite the large variability in initial strength. 3) Fusion-competent docked and primed SV\(_{\text{TS}}\) have a high \(P_{\text{fusion}}\) of \(~0.4\), consistent with the experimentally observed rapidly progressing STD during high-frequency stimulation. 4) Depending on \(P_{\text{fusion}}\) release occurs with different characteristics: 1) At very low \(P_{\text{fusion}}\) the number of newly primed SVs per ISI is primarily determined by the resting values \(k_{\text{LS,rest}}\) and \(k_{\text{TS,rest}}\) and the activity-independent unpriming rate constants \(b_1\) and \(b_2\), such that increasing \(P_{\text{stim}}\) leads to decreasing steady-state release. ii) For intermediate \(P_{\text{fusion}}\) the \(\text{Ca}^{2+}\)-dependent increases of \(k_1\) and \(k_2\) above their resting values dominate. Because individual AP-evoked \([\text{Ca}^{2+}]\) transients do not overlap, each AP causes a forward transition of a constant fraction of SVs to the respective downstream subpool, irrespective of the interval between consecutive APs. Since each AP also triggers the fusion of an almost constant fraction of \(\text{SV}_{\text{TS}}\) \((P_{\text{fusion}})\), release tends toward a steady state, which is independent of frequency. iii) At high \(P_{\text{stim}}\) additional kinetic features control release, such as an increase in \(P_{\text{fusion}}\) during trains, a speed-up of the priming process, a transiently increased occupancy of \(\text{SP}_{\text{TS}}\) contributing to release facilitation, and a frequency-dependent decline of \(\text{SP}_{\text{LS}}\) and \(\text{SP}_{\text{TS}}\) occupancies and steady-state release rate. The first three features contribute to PPF, while the fourth causes STD. For modeling these high-frequency features, we had to extend the kinetic scheme (Fig. 1 O).

Before discussing these aspects individually, we summarize the assumptions underlying the relationship between NTF analysis and model fitting.

**Assumptions underlying NTF decomposition of quantal release during eEPSC trains.** NTF-based decomposition of eEPSC trains into release components rests on the following four assumptions: 1) Contributions to release are nonnegative. 2) Release can be decomposed into distinct components, each of which represents the contribution by SVs that had been in a certain state prior to stimulation (LS, TS, or undocked). 3) Forward transition rates of SV priming and fusion strongly dominate over backward transition rates in the \(P_{\text{stim}}\) range used for NTF analysis (5–200 Hz). 4) Heterogeneity among synapses with respect to initial strength and STP characteristics is primarily due to differences in the relative abundance of \(\text{SV}_{\text{LS}}\) and \(\text{SV}_{\text{TS}}\) at rest—i.e., individual synapses are endowed with variable fractions of docked SVs equipped with a mature release machinery.

NTF decomposes release into components constrained only by nonnegativity. Therefore, a given NTF fit result should not be regarded as a unique solution but rather as one of many options consistent with both the mean eEPSC trains of all synapses examined and their variability. Provided that the above-mentioned assumptions hold, NTF analysis delivers good initial guesses for model parameters, which together with trial-and-error parameter optimization leads to a number of important conclusions discussed in the following sections.

**A simple equation for estimating \(P_{\text{fusion}}\) from low-frequency stimulation-induced STD.** When analyzing 5 to 20 Hz eEPSC trains, 1) we postulated a linear relationship between \([\text{Ca}^{2+}]\) and the priming rate constants \(k_1\) and \(k_2\) and 2) we assumed that the integral over the AP-induced \([\text{Ca}^{2+}]\) transient is constant over that \(P_{\text{stim}}\) range. This results in balanced priming, with the average SV recruitment rate and the average release rate both being proportional to \(P_{\text{stim}}\) at steady-state conditions. However, the same conclusions can be reached with much-less-specific assumptions. Any process by which an AP causes the release of a certain fraction of fusion-competent SVs and also the transition of a certain fraction of immature SVs to the fusion-competent state will reach a frequency-independent steady state. This would hold even if recruitment were nonlinearly dependent on \([\text{Ca}^{2+}]\) as long as \([\text{Ca}^{2+}]\) transients summed only little, such that each AP exerted its \(\text{Ca}^{2+}\)-dependent effect independently. This is largely the case for \(P_{\text{stim}}\leq20\) Hz. The choice of the exact time course of \([\text{Ca}^{2+}]\) transients has little influence on the outcome of simulations in this \(P_{\text{stim}}\) range, because different \([\text{Ca}^{2+}]\) transients will similarly influence release as long as they have the same time integral and decay within the ISI back to \([\text{Ca}^{2+}]\)rest (Eqs. 12–14).

The simplicity of conditions in the 5 to 20 Hz range (almost constant \(P_{\text{fusion}}\) balanced loss and gain for \(\text{SP}_{\text{LS}}\) fixed fraction of source pools transferred to downstream pools per AP) allowed us to derive a simple equation for calculating \(P_{\text{fusion}}\) from the two easily measurable quantities, \(\text{PPR}\) and \(\text{Dm}\). In this expression \((P_{\text{fusion},1} = (1 – \text{PPR}))/\left(1 – \text{Dm}\right)\); Eq. 31), the numerator recapitulates the standard use of PPR as an indicator for \(P_{\text{fusion}}\) (14, 66). The equation shows that 1 – PPR is actually a good estimate for \(P_{\text{fusion}}\) in case of complete steady-state depression (\(\text{Dm} = 0\)), provided that all simplification made in the derivation of Eq. 31 can be applied. However, SV pool depletion is usually far from complete for 5 to 20 Hz stimulation for which Eq. 31 holds. Thus, the correction by the denominator is substantial. The \(P_{\text{fusion},1}\) value obtained that way (0.43) is quite close to that derived from NTF analysis (0.39), the latter being used in all numerical simulations. Estimates for \(P_{\text{fusion},1}\) obtained according to Eq. 31 for individual synapses show less variability than corresponding \(m_1\) values, which argues against \(P_{\text{fusion},1}\) being the main source of heterogeneity in synaptic strength among synapses (Fig. 2 D–F).

**Comparison to \(P_{\text{fusion}}\) estimates obtained by "traditional" quantal analysis methods.** For a parallel priming and fusion scheme, the mean \(P_{\text{fusion}}\) is simply the weighted average of the \(P_{\text{fusion}}\) values pertaining to individual subpools of fusion-competent SVs. The mean \(P_{\text{fusion}}\) may be low, if "reluctantly releasing" SVs (9, 18, 67, 68) contribute to release. In the context of the sequential priming scheme discussed here, our analysis suggests a relatively high \(P_{\text{fusion}}\) for SV\(_{\text{TS}}\), which are the only fusion-competent SVs in resting synapses (Fig. 1 B). In contrast, "traditional" estimates often calculate \(P_{\text{fusion}}\) as the ratio of the quantal content of a single synaptic response over the size of the pool of readily releasable SVs as determined by high-frequency stimulus trains.
(69). Under the conditions described here, such stimulus trains deplete not only $SP_{75}$ but rather the sum $SP_{75} + SP_{75}$ due to the rapid LS $\rightarrow$ TS transition during high-frequency stimulation. Release probability as measured by traditional methods, therefore, is the product $p_{\text{fusion}} \cdot SP_{75}/(SP_{75} + SP_{75})$, i.e., $p_{\text{fusion}}$ times the probability of a docked SV being in the tightly docked state equipped with a mature release machinery (51). Considering a sequential priming scheme with only one fusion-competent state, $p_{\text{fusion}}$ as determined by NTFT or model fitting is a quantity strictly reflecting the SV fusion process, whereas $p_{\text{fusion}}$ as determined by traditional methods is a quantity depending on the fusion process and on the priming equilibria at rest (70, 71) (SI Appendix, Fig. S5). This insight, if applicable to a given type of synapse, has strong implications for the interpretation of perturbations of synapse function by mutagenesis or pharmacological tools: An observed change in $p_{\text{fusion}}$ is generally interpreted as a change in the fusion machinery, involving SNARE proteins, synaptotagmins, and complexins, or else reflecting changes in the microdomain $[\text{Ca}^{2+}]$ signal, which depends on $\text{Ca}^{2+}$ currents, $\text{Ca}^{2+}$ buffers, and coupling distances between voltage-gated $\text{Ca}^{2+}$ channels (VGCCs) and $\text{Ca}^{2+}$ sensors for SV fusion. According to our interpretation, an observed change in a traditional $p_{\text{fusion}}$ estimate may well reflect changes in the LS $\rightarrow$ TS equilibrium at rest, possibly involving presynaptic proteins such as Munc13, DOCC2, CAPS, and synaptotagmin7 (27, 30). A modulatory influence of second messengers ($[\text{Ca}^{2+}]$), diacylglycerol (DAG), phosphatidylinositol 4,5-bisphosphate (PIP$_2$)) may show up in traditional quantal analysis as a change in $p_{\text{fusion}}$ while NTFT analysis reports it as a shift in the state of priming at rest. More precisely, the ratio of traditional $p_{\text{fusion}}$ estimates over NTFT-derived $p_{\text{fusion}}$ estimates equals $SP_{75}/(SP_{75} + SP_{75})$. Provided that conditions can be found for a given type of synapse for which Eq. 31 applies, $p_{\text{fusion}}$ can be estimated from PPR and $D_n$ measured at a suitable frequency, and $SP_{75,\text{rest}}$ is readily calculated as the ratio $m_3/p_{\text{fusion}}$. Assuming further that pool estimates obtained by analyzing depleting high-frequency eEPSC trains represent the sum $SP_{75,\text{rest}} + SP_{75,\text{rest}}$ then both $SP_{75,\text{rest}}$ and $SP_{75,\text{rest}}$ can be determined approximately from a small number of measurements without kinetic modeling or NTFT analysis.

Synaptic facilitation and saturation of priming shape the release time course during high-frequency trains. For the $f_{\text{fusion}}$ range of 5 to 20 Hz, NTFT provides quite stringent constraints on model parameters and eEPSC trains can be modeled as independent occurring release events largely unaffected by the range of 5 to 20 Hz, $N_{\text{TF}}$ provides quite stringent constraints on the observed release time course during high-frequency trains.

Synaptic facilitation and saturation of priming shape the release time course during high-frequency trains. For the $f_{\text{fusion}}$ range of 5 to 20 Hz, each AP shifts certain constant fractions of SVs from one pool to the next downstream pool, which suggests a local, short-lived action of the effective $[\text{Ca}^{2+}]$ transients. If $\text{Ca}^{2+}$-dependent priming depends on Munc13, then a local $[\text{Ca}^{2+}]$ transient similar to that triggering SV fusion may be most relevant since Munc13 is an integral part of the AZ (79–81). On the other hand, the priming rate may saturate far below the peak $[\text{Ca}^{2+}]$ within such local domains if the respective $\text{Ca}^{2+}$ sensors have high affinity. A complete experimental value, but $m_3$ and $m_4$ are even more so. Adding the contribution of the TSL state to release leads to an adequate fit up to $m_9$ (Fig. 5B, dotted trace). However, it predicts subsequent release to rebound due to enhanced priming by the gradual buildup of $[\text{Ca}^{2+}]$ during 200 Hz trains. Only the additional implementation of $K_i$ saturation reproduces the experimentally observed release time course over the entire stimulation period (Fig. 5B, solid trace). Alternatively, the experimental data can be simulated satisfactorily by the introduction of an ERS into the model (Fig. 1 C, ii and SI Appendix, Table S2).

Comparison with parallel release models and past work on the calyx of Held synapse. Previous studies of calyx synapses using long step depolarizations or presynaptic $\text{Ca}^{2+}$ uncaging for triggering SV fusion identified two kinetically distinct release components mediated by two SV pools differentially coupled to presynaptic VGCCs (72, 73): a slowly releasing pool (SRP) and a fast-releasing pool (FRP) (68, 74, 75). The FRP accounts for the majority of SVs fusing during AP-evoked release, while the contribution of the SRP is only minor (76). The FRP may thus be regarded as the calyx–synapse equivalent of the readily releasable pool as estimated with AP trains in other synapses. Notably, $SP_{75}$ and $SP_{75}$ are not congruent with SRP and FRP but rather represent a subdivision of the FRP.

In a sequential priming scheme as used here, kinetic components may not be readily linked to specific state transitions. In contrast, it is intuitively easy to consider two kinetic components as independent contributions of two SV populations. Their intuitive tangibility explains why parallel kinetic schemes are particularly attractive. Previously, we thus described some of the STP features, based on a smaller data set, in terms of a parallel model consisting of a rapidly releasing SV subpool (called superprimed SVs) and a slowly releasing one (called normally primed SVs) (16) in analogy to studies at cultured hippocampal synapses (77). Superprimed SVs share many properties with $SV_{75}$.$SV_{75}$ Normally primed SVs in the context of a parallel kinetic scheme represent fusion-competent SVs, albeit with a low fusogenicity (16). However, in the framework of the sequential scheme, proposed here, only $SV_{75}$ are considered fusion competent. Therefore, and in view of the morphological evidence for “loose” and “tight” coupling, we consciously avoided the previous terminology.

Release mediated by $SV_{75}$ has many features in common with the release contributed by so-called preprimed SVs, postulated for glutamatergic synapses in the cortex (78). Likewise, the sequential scheme proposed for cerebellar parallel fiber–molecular–layer interneuron synapses (22, 23) is related to the sequential scheme favored here: Both assume reversible priming steps in sequence, postulate certain fractions of upstream SV pools being transferred to a downstream pool upon AP arrival, and separate contributions to release in terms of the state of SVs prior to stimulation.

The identity of the $[\text{Ca}^{2+}]$ signal regulating the priming rate. The identity of the $[\text{Ca}^{2+}]$ signal that regulates priming (effective $[\text{Ca}^{2+}]$; SI Appendix, Table S2) is not unequivocally established. For $f_{\text{fusion}}$ of 5 to 20 Hz, each AP shifts certain constant fractions of SVs from one pool to the next downstream pool, which suggests a local, short-lived action of the effective $[\text{Ca}^{2+}]$ transients. If $[\text{Ca}^{2+}]$-dependent priming depends on Munc13, then a local $[\text{Ca}^{2+}]$ transient similar to that triggering SV fusion may be most relevant since Munc13 is an integral part of the AZ (79–81). On the other hand, the priming rate may saturate far below the peak $[\text{Ca}^{2+}]$ within such local domains if the respective $\text{Ca}^{2+}$ sensors have high affinity. A complete
description of $\text{Ca}^{2+}$-dependent priming would have to consider endogenous $\text{Ca}^{2+}$ buffers, which shape the time course of both the local domain [$\text{Ca}^{2+}$]$_{\text{r}}$ (82, 83) and the global volume-averaged [$\text{Ca}^{2+}$] (64, 65, 84, 85). Here we chose to explore the influence of the [$\text{Ca}^{2+}$]$_{\text{r}}$ time course with the simplest possible model, characterized by an AP-induced increment in [$\text{Ca}^{2+}$] proportional to the $\text{Ca}^{2+}$ influx, followed by an exponential decay toward [$\text{Ca}^{2+}$]$_{\text{rest}}$ (Fig. 4C and SI Appendix, Table S2; Eq. 6).

**Sources and consequences of functional diversity among individual synapses.** Fig. 5 C and D illustrates results of two simulations during which all model parameters were identical except for $b_{\text{ax}}$ which was either increased or decreased to obtain a relative size of $SPT_{\text{TS,rest}}$ of 20% (Fig. 5 C, i and D, i) or 74% (Fig. 5 C, ii and D, ii), respectively. We then systematically varied $b_{\text{il}}$ or $b_{\text{KS,rest}}$ (Fig. 5E) and plotted the ratios $m_{\text{s}}/m_{\text{i}}$ and $m_{\text{s}}/m_{\text{1}}$ for simulated 200 Hz eEPSC trains as a function of the relative size of $SPT_{\text{TS,rest}}$ (Fig. 5 E, i). We observed negative correlations, reminiscent of the relationship between PPR and the measured $m_{1}$ (Fig. 2D). A similar negative correlation was found between the simulated steady-state depression $m_{p}/m_{1}$ and simulated $SPT_{\text{TS,rest}}$ (Fig. 5 E, ii), illustrating that the magnitude of depression strongly depends on the relative abundance of SV$_{\text{TS}}$ at rest. Fig. 5 E, ii also plots the absolute $m_{1}$ as a function of the relative size of $SPT_{\text{TS,rest}}$ for simulated 200 Hz trains. As expected, $m_{1}$ is independent of the distribution of subpool sizes at rest since the $\text{Ca}^{2+}$-dependent parameters of the priming scheme dominate at high frequencies over the small rate constants, which set the distribution of states at rest.

Recent studies emphasize variations in coupling distances between presynaptic VGCCs and docked SVs as a predominant mechanism generating differences in $p_{\text{fusion}}$ (for review, see reference 86). Our implementation of NTF analysis, on the other hand, postulates differences in the maturation state of primed SVs as the main source of intersynapse variability. It is well conceivable that a different NTF implementation, allowing several states or types of release sites to contribute to release, would come up with BFs for such states with different $p_{\text{fusion}}$ as one might expect for different coupling distances. However, our assignment of variability to differences in the maturation state of primed SVs is self-consistent in the sense that $p_{\text{fusion}}$ is independent of the distribution of SV$_{\text{TS}}$ and the Ca$^{2+}$ sensor for SV fusion may be heterogeneous.

The identification of molecular mechanisms causing functional synaptic diversity has recently attracted great attention due to the recognition of its importance for maximizing the capacity of information processing of neuronal networks (87) and of findings regarding the modulation of heterogeneity by astrocytes (88). Activity-induced changes in synaptic weight such as the redistribution of synaptic efficacy as a consequence of the induction of LTP (89) have emerged as important elements in the description of the dynamics of neuronal networks (90, 91). However, the understanding of molecular mechanisms underlying such phenomena has lagged behind. Such redistribution is a basic property of our model, controlled by the distribution of SVs between primed states at rest and not involving a change in $p_{\text{fusion}}$. In Fig. 5C, we compare simulated time courses of the release of two synapses, one with a low ratio of $SPT_{\text{TS}}/SPT_{\text{rest}}$ (Fig. 5 C, i) and the other with a high ratio of $SPT_{\text{TS}}/SPT_{\text{rest}}$ (Fig. 5 C, ii) at rest. Despite the 4.5-fold difference in initial synaptic strength ($m_{1} = 150$ vs. 684 SVs), the cumulative release during the first 20 stimul at 200 Hz is quite similar (2,198 vs. 2,695 SVs; SI Appendix, Fig. S5B). This suggests changes in the LS $\leftrightarrow$ TS equilibrium at rest and priming proteins, such as Munc13s, as the molecular basis of a redistribution of synaptic efficacy.

**Materials and Methods**

**Preparation.** Juvenile, posthatching onset (P14–16) Wistar rats of either sex were used. All experiments complied with the German Protection of Animals Act and with the guidelines for the welfare of experimental animals issued by the European Communities Council Directive. Acute brainstem slices were prepared similarly as previously described (16). See SI Appendix for details.

**Electrophysiology.** Whole-cell patch-clamp recordings were made from principal neurons of the medial nucleus of trapezoid body (MNTB) at room temperature as previously described (16). See SI Appendix for details.

**Decomposition of quantal release into distinct components using NTF.** NTF of eEPSC trains was performed similarly as previously described (51). See SI Appendix for details.

**Simulation of AP-evoked SV release and short-term plasticity.** We used a 2-step priming and fusion scheme (Fig. 1) to numerically simulate SV fusion and short-term plasticity. Details of the model and approximations for steady-state conditions that were used to constrain model parameters are given in SI Appendix.

**Data, Materials, and Software Availability.** Data sets shown in Fig. 4A and the program code required to re-create Fig. 4B and C and all other numerical simulations are available at an open repository under https://doi.org/10.5281/zenodo.6818173 (92). All other data are included in the article and/or SI Appendix.

**ACKNOWLEDGMENTS.** We thank Drs. Alain Marty and Stefan Hallermann for valuable discussions and comments on the manuscript and I. Herfort for excellent technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Cluster of Excellence EXC 2067 “Multiscale Bioimaging” (E.N.) and the DFG Collaborative Research Center 1286 “Quantitative Synaptology” (E.N.).

Author affiliations: †Emeritus Laboratory of Membrane Biophysics, Max Planck Institute for Multidisciplinary Sciences, 37077 Gottingen, Germany; ‡Department of Molecular Neurobiology, Max Planck Institute for Multidisciplinary Sciences, 37075 Gottingen, Germany; and §Cluster of Excellence “Multiscale Bioimaging”, Georg August University, 37075 Gottingen, Germany.
et al. 15. A. Losonczy, L. Zhang, R. Shigemoto, P. Somogyi, Z. Nusser, Cell type dependence and variability in the short-term plasticity of EPSCs in identified mouse hippocampal interneurons. J. Physiol. 542, 193–210 (2002).

H. Taschenberger, A. Wonaler, E. Neher, Superpriming of synaptic vesicles as a common basis for increased vesicular availability and modulation of synaptic strength. Proc. Natl. Acad. Sci. U.S.A. 113, E4548–E4557 (2016).

J. Tommienhauer, R. Schneppenberg, A. Zippelius, E. Neher, Heterogeneous presynaptic release probabilities. Functional relevance for short-term plasticity. Biophys. J. 84, 1566–1579(2003).

F. Dossi et al., Fragments of neurotransmitter recycling facilitate short-term modulation of heterogeneous pools of synaptic vesicles shapes presynaptic plasticity. eLife 6, e28935 (2017).

S. Halamanen et al., Bassoon speeds vesicle reloading at a central excitatory synapse. Neuron 68, 710–720 (2010).

B. Pan et al., Z. Züker, A general model of synaptic transmission and short-term plasticity. Neuron 52, 539–554 (2006).

G. S. Fitzgerald et al., A new kinetic framework for synaptic vesicle trafficking tested in synapsin knockouts. J. Neurosci. 31, 11563–11577 (2011).

T. Miki, Y. Nakamura, M. Galazka, E. Neher, A. Marty, Two-component latency distributions indicate two-vesicle-step unimolecular fusion at a simple glutamatergic synapse. Nat. Commun. 9, 3943 (2018).

V. Tran, T. Miki, A. Marty, Three small vesicular pools in sequence govern spontaneous vesicle response dynamics during active action potentials. Proc. Natl. Acad. Sci. U.S.A. 119, e2144691119 (2022).

A. Martínez-Valencia, G. Ramírez-Santiago, F. F. De Miguel, Dynamics of neuromuscular transmission reproduced by calcium-dependent and reversible serial transitions in the vesicle fusion complex. Front. Synaptic Neurosci. 13, 785361 (2022).

J. S. Dittman, Unc13: A multifunctional synaptic marvel. Neuron 100, 808–816 (2018).

T. Xu, S. Hallermann, J. D. Goutman, O. Kochubey, R. Schneggenburger, Interaction between facilitation and depression at a large CNS synapse reveals mechanisms of short-term plasticity. J. Neurosci. 30, 2007–2016 (2010).

J. Turecek et al., Variable vesicle priming shapes presynaptic frequency response mechanisms at glutamatergic synapses. J. Neurosci. 27, 2261–2271 (2007).

J. Turecek, S. L. Jackman, W. G. Regehr, Synaptotagmin 7 confers frequency invariance onto synaptic strength and depression. Neuron 599, 5301–5327 (2018).

S. Hallermann et al., Inhibition of SNARE complex assembly differentially affects kinetic components of synaptic vesicle priming. Cell Rep. 33, 3219–3236 (2021).

W. G. Regehr, Synaptic specializations support frequency-independent neurotransmitter release. Neuron 66, 216–229 (2009).

F. H. Yu et al., Blocking endocytosis enhances short-term synaptic depression under conditions of normal availability of vesicles. Neuron 80, 343–349 (2013).

J. Turecek et al., Fast neurotransmitter release regulated by the endocytic scaffold interactor. Proc. Natl. Acad. Sci. U.S.A. 110, 8266–8271 (2013).

M. Müller, J. D. Goutman, O. Kochubey, R. Schneppenberg, Interference between facilitation and depression at a large CNS synapse reveals mechanisms of short-term plasticity. J. Neurosci. 30, 2007–2016 (2010).

J. Turecek et al., Munc13-1 is a Ca2+- phospholipid-dependent vesicle priming hub that shapes synaptic short-term plasticity and enables sustained neurotransmission. Neuron 109, 3990–4000 (2020).

K. H. Lin et al., Inhibition of Munc13-1 by a negative regulatory module reduces presynaptic calcium loss and enhances release probability. Neuron 87, 5329–5336 (2017).

S. Hallermann et al., Diverse types of synaptic vesicles promote distinct presynaptic functions. Neuron 100, 808–816 (2018).

T. Xu et al., Inhibition of SNARE complex assembly differentially affects kinetic components of synaptic vesicle priming. Cell Rep. 13, 1277–1288 (2015).

J. He et al., Munc13-1 and Munc18-1 together prevent NSF-dependent de-priming of synaptic vesicles. Nat. Commun. 8, 15917 (2017).

T. Miki et al., Actin- and myosin-dependent loading of preysynaptic docking sites prior to exocytosis. Neuron 91, 808–816 (2016).

T. Sakaba, Two Ca2+-dependent steps control synaptic vesicle fusion and replenishment at the cerebellar basket cell terminal. Neuron 57, 406–419 (2008).

K. R. Köberlasmeyer et al., Rapid regulation of vesicle priming explains synaptic facilitation despite heterogeneous vesicle Ca2+-channel densities. eLife 9, e51022 (2020).

D. Zenisek, J. A. Steyer, W. Almers, Transport, capture and exocytosis of single synaptic vesicles at active zones. Nature 406, 849–854 (2000).

V. N. Murthy, C. F. Stevens, Reversal of synaptic vesicle docking at central synapses. Nat. Neurosci. 2, 503–507 (1999).

S. Chang, T. Trimbuch, C. Rosenmund, Synaptotagmin 1 drives synovial Ca2+-triggered fusion by C-terminal domain-mediated synaptic vesicle-membrane attachment. Nat. Neurosci. 21, 33–40 (2018).

G. F. Kuchik et al., Vesicles transcytoplasmic dock to refill release sites. Nat. Neurosci. 23, 1333–1338 (2020).

E. Neher, N. Böse, Dynamically primed synaptic vesicle states: Key to understand synaptic short-term plasticity. Neuron 100, 1283–1291 (2018).

J. H. Jung, L. M. Kirk, J. N. Bourne, K. M. Harris, Shortened tethering filaments stabilize presynaptic vesicles in support of elevated release probability during LTP in rat hippocampus. Proc. Natl. Acad. Sci. U.S.A. 112, 20018653118 (2021).

R. Martin et al., β-adrenergic receptors/epac signaling increases the size of the readily releasable pool of synaptic vesicles required for parallel fiber LTP. J. Neurosci. 40, 8604–8617 (2020).

M. E. Balsinde-Báscones et al., Cryo-electron tomography reveals a critical role of RIM in vesicular synaptic tethering. J. Cell Biol. 201, 725–740 (2013).

C. Iimori et al., The morphological and molecular nature of nerve ending synaptic vesicle priming at presynaptic active zones. Neuron 84, 416–431 (2014).

N. Schierr et al., Articulation: Kinetics of both synchronous and asynchronous quantal release during trains of action potential-evoked EPSCs at the rat calyx of Held. J. Physiol. 575, 361–381 (2007).

C. H. Yang, W. K. Ho, S. H. Lee, Postnatal maturation of glutamate clearance and release kinetics at the rat and mouse calyx of Held synapses. Synapse 75, 62225 (2021).

H. Taschenberger, R. M. Leão, K. C. Rowland, G. A. Spirou, H. von Gersdorff, Optimizing synaptic architecture and efficiency for high-frequency transmission. Neuron 36, 1127–1143 (2002).

H. Taschenberger, V. Schues, E. Neher, Release kinetics, quantal parameters and their modulation during short-term depression in developing a synapse in rat CNS. J. Physiol. 565, 513–537 (2005).

S. Ichikawa, T. Takahashi, Developmental regulation of transmitter release at the calyx of Held in rat auditory brainstem. J. Physiol. 534, 861–871 (2007).

M. Müller, F. Felmy, R. Schneppenberg, A limited contribution of Ca2+ current facilitation to paired-pulse facilitation of transmitter release at the rat calyx of Held. J. Physiol. 586, 5503–5520 (2008).

R. Z. Lucker, Changes in the statistics of transmitter release during facilitation. J. Physiol. 229, 797–810 (1973).

A. Cichoki, R. Zündel, A. H. Phan, S. Amam, Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multivariate Data Analysis and Blind Source Separation (Wiley, Chichester, UK, 2009).

E. Neher et al., H. Taschenberger, Developmental function as a tool to distinguish between synaptic vesicles in different functional states. Neuroscience 458, 182–202 (2021).

J. Turecek, S. L. Jackman, W. G. Regehr, Synaptic specializations support frequency-independent Purkinje cell output from the cerebellar cortex. Cell Rep. 17, 3256–3268 (2016).

J. Turecek, S. L. Jackman, W. G. Regehr, Synaptotagmin 7 confers frequency invariance onto specialized desynapticing synapses. Nature 551, 503–506 (2017).

H. von Gersdorff, R. Schneppenberg, S. Wess, E. Neher, Presynaptic depression at a calyx synapse: The small contribution of metabotropic glutamate receptors. J. Neurosci. 17, E1813–E1816 (1997).