Maintenance of synaptic plasticity by negative-feedback of synaptic protein elimination: Dynamic modeling of KIBRA-PKMζ interactions in LTP and memory

Harel Z. Shouval[1,2], Rafael E. Flores-Obando[3], and Todd C. Sacktor[3,4]

[1]Department of Neurobiology and Anatomy, University of Texas Medical School, Houston, TX 77030, USA
[2]Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005, USA
[3]Department of Physiology and Pharmacology, SUNY Downstate Health Sciences University, Brooklyn, NY. 11203. USA.
[4]Department of Physiology, Pharmacology, Anesthesiology, and Neurology, SUNY Downstate Health Sciences University, Brooklyn, NY. 11203. USA.

September 25, 2024

Abbreviated title: Maintenance by negative-feedback

Corresponding author: Harel Z. Shouval, 6431 Fannin St. MSB 7.264, Houston, TX, 77030, harel.shouval@uth.tmc.edu

Acknowledgements: This work is supported by NIH funding RO1 DA034979 (H.Z.S., T.C.S.), RO1 MH53576 (T.C.S.), 2R37 MH057068 (T.C.S.), R01 NS105472 (T.C.S), R01 NS108190 (T.C.S), and MH115304 (T.C.S) .

Abstract

Activity-dependent modifications of synaptic efficacies are a cellular substrate of learning and memory. Current theories propose that the long-term maintenance of synaptic efficacies and memory is accomplished via a positive-feedback loop at the level of production of a protein species or a protein state. Here we propose a qualitatively different theoretical framework based on negative-feedback at the level of protein elimination. This theory is motivated by recent experimental findings regarding the binding of $PKM\zeta$ and KIBRA, two synaptic proteins involved in maintenance of memory, and on how this binding affects the proteins' degradation. We demonstrate this theoretical framework with two different models, a simple abstract model to explore generic features of such a process, and an experimentally motivated phenomenological model. The results of these models are qualitatively consistent with existing data, and generate novel predictions that could be experimentally tested to further validate or reject the negative-feedback theory.

1 Introduction

We all have memories that date back to our youth; we remember the house we lived in at age four; we remember a favorite or least favorite schoolteacher. These memories are believed to be encoded primarily via long-lasting synaptic plasticity that persistently modifies specific neuronal circuits in our brain (Martin et al., 2000; Whitlock et al., 2006; Yang et al., 2016; Langille and Brown, 2018), and at the molecular level this is achieved by altering the numbers or conformational states of certain proteins in these synapses (Bear and Malenka, 1994; Nicoll, 2017). However, this putative cellular basis of memory relies on proteins that typically have lifetimes orders of magnitude shorter than the memory. Here exactly lies a fundamental problem of long-term memory and synaptic plasticity: How can memories be stored for a human lifetime on the basis of proteins that are continuously degrading (Crick, 1984; Lisman, 1985). Moreover, since synaptic plasticity is a synaptic rather than a whole cell event, what affects the lifetimes of individual synaptic efficacies and hence memory is likely not the lifetime of a protein in the cell, but the time that a protein typically dwells in a synapse. This variable, the synaptic lifetime, may well be significantly shorter than the protein lifetime due to diffusion and trafficking of proteins.

To account for how such an inherent instability of memory can be overcome, different theories have been proposed (Lisman, 1985; Lisman and Zhabotinsky, 2001; Bhalla and Iyengar, 1999; Aslam and Shouval, 2012; Jalil et al., 2015). Most theoretical models assume a common mathematical structure based on a positive-feedback mechanism that can generate bi- or multi-stability. The biophysical instantiation of the feedback varies from model to model. This feedback could be based on on single species feedback loops such as autophosphorylation (Lisman, 1985; Lisman and Zhabotinsky, 2001), or more complex multi-molecule post-translational feedback loops (Bhalla and Iyengar, 1999), or, alternatively, on positive feedback loops operating at the level of the regu-

lation of translation (Aslam et al., 2009; Jalil et al., 2015; Richter and Klann, 2009; Ogasawara and Kawato, 2010). Note that despite the different biophysical substrates of these theories, their mathematical structure is similar.

One prominent candidate for a molecule that is necessary for the maintenance of late-phase long-term synaptic plasticity (L-LTP) and long-term memory is the constitutively active form of atypical protein kinase C, $PKM\zeta$ (Sacktor et al., 1993; Ling et al., 2002; Pastalkova et al., 2006). The levels of $PKM\zeta$ persistently increase after L-LTP and memory formation (Hsieh et al., 2021), and blocking its activity reverses established memory (Pastalkova et al., 2006). How this persistent increase is maintained despite protein turnover is achieved is still unclear. Previous models (Ogasawara and Kawato, 2010; Jalil et al., 2015) have postulated that this might occur due to a feedback loop based on enhanced translation in synaptic spines. The results of these models seem consistent with experiments; however their fundamental assumption has not been proven experimentally and the mechanism for such increased translation have not been elucidated.

Recent results have shown that maintenance of plasticity and memory depends on the binding between $PKM\zeta$ to another synaptic molecule, KIdney BRAin protein (KIBRA) (Tsokas et al., 2024). During L-LTP the synaptic concentration of $PKM\zeta$ bound to KIBRA increases selectively in activated pathways (Tsokas et al., 2024). Agents that specifically interfere with the binding of $PKM\zeta$ and KIBRA can reverse established L-LTP and long-term memory as late as a month after memory formation (Tsokas et al., 2024). In contrast, these inhibitors do not affect baseline synaptic transmission and do not affect the binding of other kinases such as other atypical protein kinase C molecules and Ca^{2+}/calmodulin-dependent protein kinase, II (CaMKII) to KIBRA (Tsokas et al., 2024). In addition, previous results show that the binding of $PKM\zeta$ to KIBRA shields $PKM\zeta$ from proteasomal degradation (Vogt-Eisele et al., 2014). Together, these results suggest that the increase of $PKM\zeta$ magnitude in synapses during long-term plasticity and memory arises from a decrease in protein turnover and possibly decreased diffusion rather than from an increase in synthesis. There are currently no mathematical theories of memory maintenance based on the conditional decrease of protein elimination.

Motivated by these results we propose a novel theory based on negative feedback of protein elimination. In the context of this general theoretical framework, the concept of elimination incorporates several molecular processes, including classical protein degradation, but also changes of the conformational state of a protein from an active to an inactive state, as well as the diffusion or trafficking of proteins out of the synaptic compartment. We explore here two specific instantiations of the negative-feedback theory, a simple single species model to explore the principles of maintenance by negative feedback and a more complex phenomenological model motivated by the binding of $PKM\zeta$ and KIBRA (Vogt-Eisele et al., 2014; Tsokas et al., 2024). Using these models we will outline the general properties of the negative-feedback theory, examine under what general conditions it can produce stability, and propose experimentally testable predictions.

2 Results

This paper aims to explain how maintenance can be achieved by negative feedback of protein elimination, and this is the main focus of the results section. Before turning to the analysis of the negative-feedback theory we will begin by examining a simplified synaptic state model (section 2.1), and explain how positive-feedback theories operate in this framework (section 2.2). This section is not novel work but a simple generic description of most previous models that try to explain maintenance of synaptic efficacies. We provide this section in order to contrast these theories with the negative-feedback framework. Using the same simple 1D dynamical structure we will develop a simple instantiation of the negative-feedback theory (section 2.3). We will show that this model can result in bi-stability and even generate a continuous attractor. We also outline some generic testable predictions of the theory. Finally in section 2.4 we describe a phenomenological model to account for the negative-feedback of degradation arising from interactions between $PKM\zeta$ and KIBRA, as observed experimentally (Vogt-Eisele et al., 2014; Tsokas et al., 2024). This phenomenological model is no longer a 1D model, as it is described by 4 coupled dynamical equations, but it generates a similar qualitative behavior as the simpler 1D model of negative-feedback. However, it suggests a possible mechanistic basis for the origin of the negative feedback, and it allows us to simulate the effect of inhibiting the binding of $PKM\zeta$ and KIBRA, which has been shown experimentally to reverse L-LTP and long-term memory (Tsokas et al., 2024).

2.1 The general mathematical structure of long-term synaptic plasticity in 1D

Synaptic plasticity is a complex high-dimensional process, but here for simplicity we assume a simple 1D model. In some cases the dimensionality of such high-dimensional dynamical systems can be rigorously reduced (Aslam and Shouval, 2012; Agarwal et al., 2012; Gabbiani and Cox, 2017); here we simply assume this can be done. In this 1D system the variable P represents the species of interest. This species could be a concentration of a protein, the concentration of a specific state of a protein, for example phosphorylated protein, or it can even represent a complex of bound proteins. In terms of this 1D system a dynamical equation that governs the concentration of this species (P) has the form:

$$\frac{dP}{dt} = Source\,(P) - Sink\,(P)\,. \tag{1}$$

Where the source (production) term includes everything that brings P into the compartment, including diffusion, trafficking, protein synthesis, and post-translational modifications such as phosphorylation, which produce this form of the protein within this compartment, or the binding reaction of different proteins to a compound. The sink term includes every mechanism that eliminates P from the compartment, including diffusion,

protein turnover, and processes such as dephosphorylation or unbinding. If dimensionality reduction techniques are used, the variable P could also be some transformed variable representing a combination of some other elementary variables. Note that both the sink and source terms depend on P, itself, as both of the positive-feedback and the negative-feedback theories require different forms of this dependence on P.

Maintenance theories are typically based on equations that have several stable fixed points, and these stable fixed points represent stable states of synaptic efficacy that persist despite ongoing protein degradation and diffusion. In order to have several fixed points the sink and source terms must cross several times (Fig. 1A), and for this to happen at least one of these two terms must be nonlinear. Most current theories, which depend on positive feedback in the source term, assume that the interesting non-linear part resides in the source term, and that the sink term is linear. The non-linearity of the source term is assumed to arise from processes such as autophosphorylation or the proteins control of its own translation rate within the compartment. In the negative-feedback framework, the interesting, non-liner term is in the sink term, and will be due to negative-feedback of one or more of the processes involved in the actual elimination of proteins from the compartment, which includes protein degradation, diffusion, and trafficking.

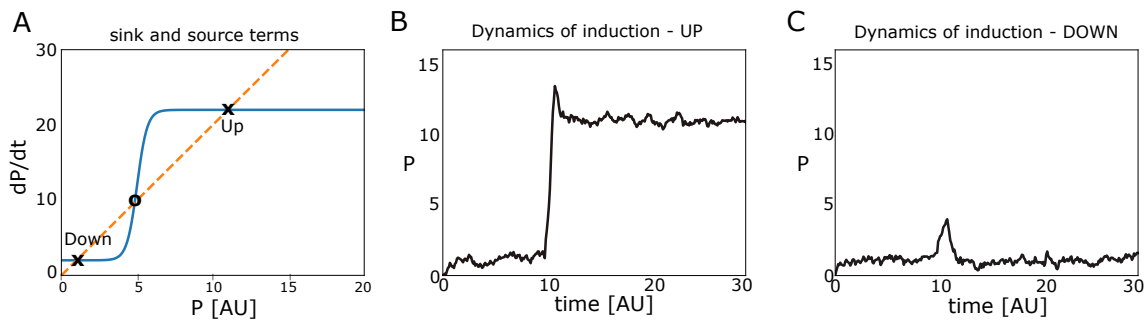


Figure 1: Bi-stability with positive feedback in the source term. A. Sink term (dashed red line) and source term (blue line). These lines cross at three points, two of these are stable fixed-points (x) and one is unstable (o). B. Dynamics of L-LTP induction, triggered by a large pulse of a current of P starting at time T=10 [AU]. Here, the system successfully shift from the DOWN to the UP state (dynamics include noise). C. With a smaller input pulse there is only a transient change and the system returns to the DOWN state.

2.2 Positive-feedback theories of maintenance

In positive-feedback theories, the source term is more complex and is the mechanistic origin of synaptic stability, while the sink term is usually assumed to be linear of the form: $Sink\,(P) = \lambda P$, where λ is the rate constant. More complex, yet monotonic sink terms typically do not alter the qualitative behavior of the system. When the source term has a nonlinear shape, for example, like the source term in figure **??**A, then the sink and source

terms cross several times. When the source and sink terms cross, the left hand side of equation 1 is zero, and hence $dP/dt = 0$, which is the definition of a fixed point. In the example shown in figure 1 they cross 3 times, two of these crossings denoted by UP and DOWN are stable fixed points, which correspond to synaptic efficacy in its basal state (DOWN) and after L-LTP (UP). The intermediate state is not stable which means that if the synapse is in that state, small fluctuations will drive the system away from that state towards either the DOWN or UP states.

In Figure 1B we show the simulated induction of LTP in such a system, using a short duration of an additional input current of P. If this current is sufficient L-LTP is induced and maintained (Fig. 1B). If the current is too small there is a transient increase in P but there is no maintenance (Fig. 1C). These simulations include a noise term to show that the stable fixed-points are indeed preserved when noise is added (Fig. 1). Noise here was added phenomenologically, and not using a rigorous approach (Gillespie, 2002), as it is non-trivial to add noise correctly to a model with phenomenological rather than elementary kinetics (Agarwal et al., 2012). This approach to adding noise, which is used to understand the qualitative rather than quantitative effects of noise, is used throughout this paper. Note, that it is essential that the non-linearity of the source term be steep enough, or ultrasensitive to obtain bistability. If the source term is not ultrasensitive then the system has only a single fixed point.

The origin of the non-linearity (Fig. 1A) in such systems can arise from various types of molecular interactions, for example auto-phosphorylation (Lisman, 1985), or the up regulation of the translation of new protein in a protein dependent manner (Aslam and Shouval, 2012; Jalil et al., 2015). These different molecular and cellular mechanisms are mathematically similar as they are all based on positive feedback. The mathematical formulation of these different biophysical processes should be described by additional variables, and therefore the complete theory is faithfully characterized by a higher dimensional system. However, within the crucial dimension described by equation 1 the implication of the positive feedback is a steep non-linearity. As shown, positive feedback theories can be bistable, however whether such systems indeed manifest bi-stability depends on the other variables in the original higher dimensional system, and on the systems parameters.

For one simple example of a positive feedback theory, proposed by Lisman in 1985 (Lisman, 1985), it is possible to reduce the higher dimensional system to a 1D system, and exactly solve for the fixed-points of the system. This was done approximately in Lisman's 1985 original paper, and we subsequently derived a precise reduction to a 1D equation (Agarwal et al., 2012). This is one example that demonstrates how it is analytically possible to reduce a higher dimensional system to a 1D system, in which the non-linear interactions can generate bi-stability. More complex and realistic models have been developed using the autophosphorylation of CaMKII and the details of the CaMKII holoenzyme (Lisman and Zhabotinsky, 2001; Miller et al., 2005). Although such models are much more complex, their steady states can sometimes be calculated analytically (Gabbiani and Cox, 2017), and their qualitative behavior may be captured by the 1D model. Another source of a

positive feedback loop can arise from the regulation of translation within synaptic spines, previous models have modeled the regulation of the synthesis of CaMKII via the protein CPEB1 (Aslam et al., 2009; Aslam and Shouval, 2012) or the regulation of the synthesis of $PKM\zeta$ (Jalil et al., 2015). Although the biological mechanism is totally different, these different models share a similar mathematical structure.

Positive feedback models can also generate multi-stability as we have previously shown (Jalil et al., 2015) if the source term has a more complex shape due to multiple feedback mechanisms. Such multi-stability has been both modeled and synthetically constructed in whole mammalian cells at the level of transcription (Zhu et al., 2022).

2.3 Stability due to negative feedback of protein turnover: abstract model.

The negative-feedback theory requires a non-linear sink term and for simplicity one might assume a constant source term. Here we explore a very simple instantiation of this theory in which we will assume a single protein P with a variable sink term and a single synaptic compartment. We will explore below a more complex model (section 2.4), but first we will present a very simple model to illustrate the concept.

Assume for now that there is a single protein that can undergo a steep transition between two states. In one of these states the degradation rate (λ_1) is fast; in the other state the degradation rate (λ_2) is slower ($\lambda_1 > \lambda_2$). The variables P_1 nd P_2 denote the concentration of protein in states 1 and 2, respectively, and the total protein concentration is $P = P_1 + P_2$. We define a state function $f(P, \Theta_P)$ that determines the fraction of protein in state 2, and assume that it depends on the total concentration of the protein P, and a set of parameters Θ_P. According to these definitions, $P_2 = f(P, \theta_P) \cdot P$, and $P_1 = (1 - f(P, \theta_P)) \cdot P$. Under these assumptions the dynamical equation would take the form:

$$\frac{dP}{dt} = I_p - (\lambda_1 (1 - f) + \lambda_2 f) \cdot P. \tag{2}$$

Where I_P is the source term, which for simplicity assume is constant, except during the induction phase, in which it is transiently increased. We have omitted the arguments of the state function f, for simplicity. The function f should be itself modeled as a dynamical process as well, but let us assume we can ignore these dynamics (maybe they are much faster) and that they could be replaced by a function that accounts for their steady-state behavior. For such a 1D system, fixed-points are formed when the source term crosses the sink term. In Figure 2A-C we present these two curves for several different choices of f. We use a sigmoidal state function of the form $f(x, \beta, \theta) = 1/(1 + exp(-\beta(x - \theta)))$ with two parameters, a slope denoted as β and half max denoted as θ. In the limit $\beta \to \infty$ the sigmoid becomes a step function.

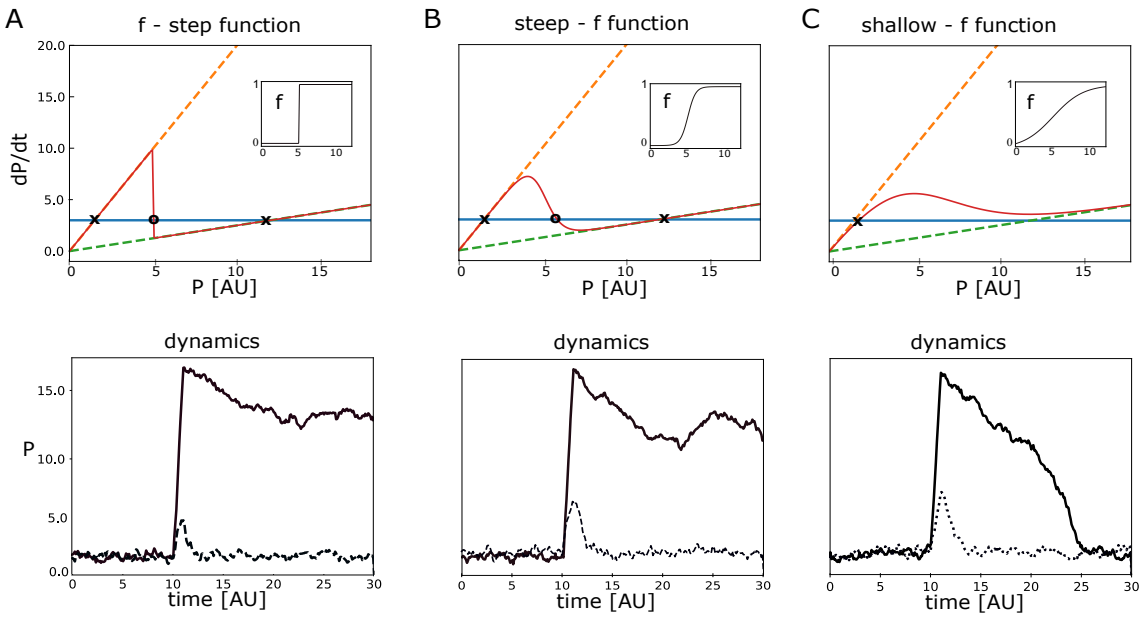


Figure 2: **Bistability with negative feedback in the sink term.** **A**. Bistability in the case of a state function (f) being a step-function ($\beta \to \infty,\ \theta = 5$). Top - A constant source (blue line) term, and two different sink terms with different values of λ (dashed orange and green lines). The resulting effective sink term (from eq. 2) given a step state function f (red line). The shape of the state function used is shown in the inset. The effective sink term crosses the source term at 3 points. Stable fixed-points are marked with **x**; unstable fixed-points with **o**. Botton - simulations of the induction of L-LTP with an additional noise term. Solid line - large initial stimulus, Dashed line - sub-threshold initial stimulus. **B.** Bistability with f a steep sigmoid function ($\beta = 2,\ \theta = 5$). Top- The resulting effective sink (eq. 2) term given a steep sigmoidal state function f. The shape function used is shown in the inset. Bottom - as in A, for the steep sigmoid. **C.** Mono-stability is obtained when f is not steep enough ($\beta = 0.5,\ \theta = 5$). Top - shows source and effective sink terms, as in A and B. There is only one intersection between the source and sink terms. Bottom- simulations always converge to the same fixed point; the system is mono-stable.

The first two examples, in which the state-function is a step-function (Fig. 2A) and a steep sigmoid (Fig. 2B) exhibit bi-stability. When the state function is not sufficiently steep (Fig. 2C), the system has only a single stable fixed-point. The lower portion of subplots in Figure 2 show simulations in which we emulate the induction of L-LTP with a transient increase in protein synthesis (I_P). When the system is bistable L-LTP can be induced (Fig. 2A,B - bottom), when it is monostable only a transient increase in P is obtained (Fig 2C-bottom). These simulations include a noise term, responsible for the fluctuations observed in the dynamics. With the given magnitude, these fluctuations, do not destabilize the fixed-points; demonstrating that they are indeed stable. In this bistable system there are two stable fixed points denoted as the DOWN and UP state. We denote the total protein levels in the DOWN and UP states as P^D, P^U respectively. The magnitude of the two states of the protein in the DOWN and UP fixed points are denoted as: $P_1^D, P_2^D, P_1^U, P_2^U$, respectively. The total protein concentration is each state is $P^D = P_1^D + P_2^D$ and $P^U = P_1^U + P_2^U$.

In the limiting case when f is a step function ($\beta \to \infty$), there is a simple relationship between the levels of the total protein in the UP and DOWN states and the degradation rates, of the form:

$$P^U/P^D = \lambda_1/\lambda_2. \tag{3}$$

This relationship is also approximately true for a large range of $f-$function parameters that are steep enough to maintain bi-stability, and can therefore be seen as a prediction of the model. Subsequently we will test if this prediction generalizes to more complex realizations of the theory.

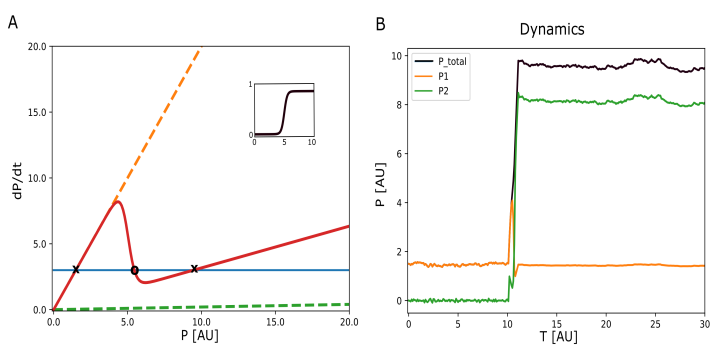


Figure 3: **Bistability with negative feedback with a small change in P_1 between DOWN and UP states. A**. Sink and Source terms when the transition function (f) saturates below 1 (inset), specifically $f_{max} = 0.85$. The ratio of $\lambda_1/\lambda_2 = 100$. There are two stable fixed points (x) and one unstable fixed point (o). **B.** Dynamics of the saturating model. Induction from $10 - 11$ in the arbitrary time units. The level of P_2 increases significantly, but P_1 stays relatively stable.

A necessary condition for obtaining bi-stability in this model is that the total amount of P_1 in the UP state (P_1^U) must be lower than in the DOWN state (P_1^D). This is a necessary condition because in the negative feedback model the production of protein is constant, so all that controls the level of proteins in the two states is the turnover. In the examples given above, the range of the state function (f) is from 0 to 1. With this assumption, in the DOWN state almost all the protein is in the P_1 form, and in the DOWN state almost all protein is in the P_2 form.

This almost all-or-non scenario is not essential for bi-stability. Consider a different f function that saturates at a value $f_{max} < 1$ for example: $f(x, \beta, \theta) = f_{max}/\left(1 + exp(-\beta(x - \theta))\right)$. We will now assume that in the DOWN state the value of $f \approx 0$ and in the UP state it is $f \approx f_{max}$; an assumption that is approximately correct for a sharp f function. In such a case the protein level in the DOWN state is still approximately $P_1^D = I_p/\lambda_1$, and $P_2^D \approx 0$. In the UP state the concentrations are approximately:

$$P_1^U = \frac{I_p(1 - f_{max})}{\lambda_1 + (\lambda_2 - \lambda_1)f_{max}} \tag{4}$$

$$P_2^U = \frac{I_p f_{max}}{\lambda_1 + (\lambda_2 - \lambda_1)f_{max}} \tag{5}$$

In order to obtain stability in such conditions it is also useful to have much lower degradation of the P_2 protein, that is $\lambda_2 << \lambda_1$. Using this condition we find that:

$$\frac{P_1^U}{P_1^D} \approx 1 - \frac{\lambda_2}{\lambda_1} \cdot \frac{f_{max}}{1 - f_{max}} \tag{6}$$

This implies that if indeed $\lambda_2 << \lambda_1$, the total amount of P_1 in the UP state could be quite close to its amount in the DOWN state, though still necessarily lower. An example of this can be seen in Figure 3. Here the ratio $\lambda_1/\lambda_2 = 100$, and $f_{max} = 0.85$. According to the theoretical approximation above (eq: 6) this yields a ratio of 0.943. In the simulations $< P_1^U > / < P_1^D >= 0.947$ (Fig. 3), when averaged over the periods in each different states state.

The form of state function (f) determines if the system mono- or bi-stable. One could imagine other functional forms of f for which the system might be multi-stable. In the extereme case this system can also become a continuous attractor over a limited range and for a very specific state function. A form of the state function that can generate such a continuous attractor can simply be found by setting equation 2 to zero, and solving for f. This produces the equation:

$$f = \frac{\lambda_1 P - I_P}{(\lambda_1 - \lambda_2)\, P} \tag{7}$$

This equation is valid for P such that $f \geq 0$; otherwise it is set to zero.

With this choice of the state-function all levels of P in the range $I_p/\lambda_1 \leq P \leq I_p/\lambda_2$ are stable; this is the range of continuous stability. In Figure 4A we show the source and sink

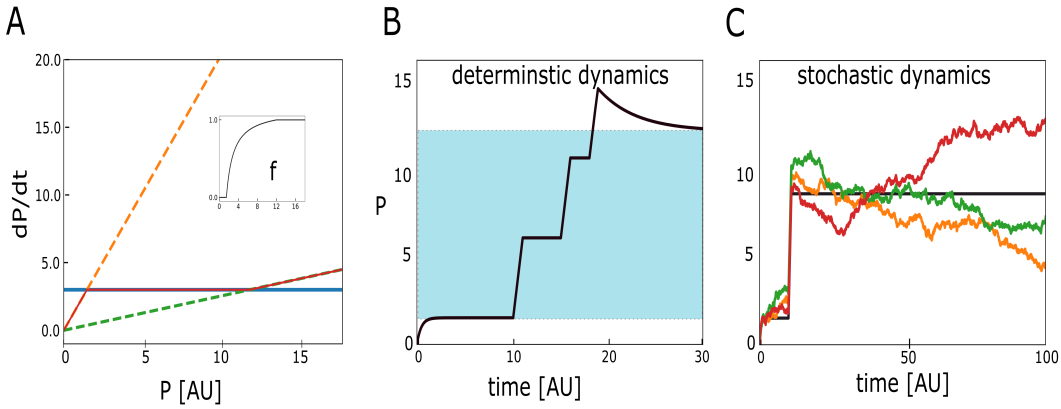


Figure 4: A continuous attractor resulting from well-tuned negative feedback. A. The source and sink terms, labeled as in previous figures. The sink and source terms have a substantial overlap by construction, as determined by using the state-function in equation 7 (see inset). B. Simulating the continuous attractor model. Initially the model settles at the lowest possible persistent state where $P = I_P/\lambda_1$. A set of three consecutive induction stimuli cause jumps to new persistent levels of P. The last stimulus causes a large jump which decays back to the maximal allowed persistent level of I_P/λ_2. The light blue shaded area represents the range of continuous stability. C. Stochastic simulations of the system. The deterministic simulation (black) exhibits persistent activation. Three different stochastic simulations (orange, red, green) show significant fluctuations within the "stable" region. In simulations the parameters are: $I_P = 3, \lambda_1 = 2, \lambda_2 = 0.25$, all in arbitrary units [AU]. The added noise in C has mean 0 and a standard deviation of 0.25 in [AU].

terms given this form of the state-function, and in the inset the form of f is displayed. In Figure 4B we show the results of several different consecutive induction episodes triggered by a transient increase in I_P. As above. Every transient causes a change in the steady state of the system, as loong P stays within the allowed range. This continuous attractor is an integrator as it keeps a stable record of the integral of the inputs into the source term. With no noise as in Figure 4B, these steady states are stable. However, if noise is added to these simulations (Fig. 4C) there are fluctuations of the state within the allowed range. The magnitude of the fluctuations, as well as the velocity of the drift, increase with the noise level. Such fluctuations within the stable range are an inherent property of continuous attractors. Note that the time axis here is labeled with arbitrary units ([AU]), which means that we do not know the time scale of the drift along the continuous attractor. With such a simple and abstract model there is no way to set real time units. With more complex models in which actual molecular reactions such as synthesis, diffusion, degradation and binding take place, it would be possible to set physical time units, but only once these biophysical parameters are estimated.

In the model of maintenance based on negative feedback of the sink term, the non-

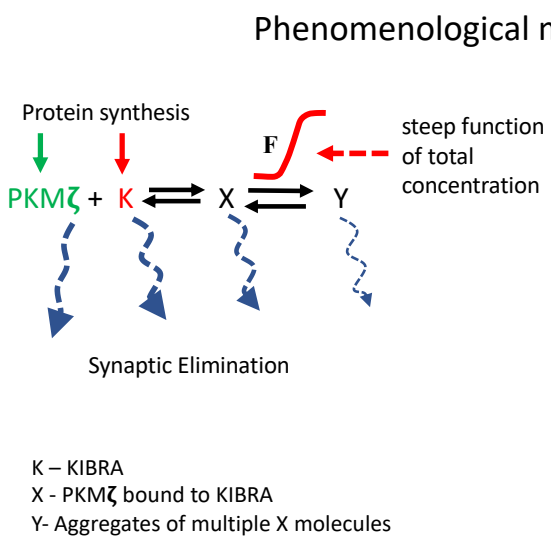
linearity responsible for bi-stability is in the sink term, rather than in the source term. A necessary condition for obtaining stability is the steep (ultrasensitive) state function. Here we assume it exists, but do not explain its mechanistic origin. In section 2.4 we propose a more detailed model motivated by experimental evidence regarding interactions between specific molecules. That model will no longer be a 1D model, though it still belongs to the negative-feedback theoretical framework. Many of the major qualitative features of the abstract model, including the dependence on the different degradation rates and the critical importance of the ultrasensitive process, are maintained.

2.4 A phenomenological model based on $PKM\zeta$ and KIBRA binding.

Recent results have shown that L-LTP and long-term memory depend of the binding between two synaptic proteins, $PKM\zeta$ and KIBRA (Tsokas et al., 2024). After L-LTP the concentration of the bound proteins increases selectively in synapses. Molecules that selectively inhibit the binding of these proteins can reverse L-LTP without affecting basal synaptic transmission. These selective inhibitors can also inhibit memories up to a month after they were established. It is also known that the binding on $PKM\zeta$ to KIBRA shields $PKM\zeta$ from proteasomal degradation (Vogt-Eisele et al., 2014; Tsokas et al., 2024).

Recently we have also shown using an expression system composed of HEK293 cells that the co expression of $PKM\zeta$ and KIBRA results in apparent droplets in which there is a a high density of both proteins(Tsokas et al., 2024). If each of them is expressed alone, such droplets do not appear. If either of two inhibitors that prevent the binding of $PKM\zeta$ and KIBRA (ζ-stat or K-ZAP) are applied these droplets do not appear, indicating that binding between these proteins is necessary for the formation of these droplets. Additionally, if these cells are photo-bleached there is a partial recovery of fluorescence within a few minutes (results no shown). Crucially, we have reanalyzed these results, as shown in the supplementary Figure, and found that the steady state concentration of both $PKM\zeta$ and KIBRA increases when they are co-expressed and able to bind. If their binding is prevented using ζ-stat or K-ZAP(Tsokas et al., 2024), their steady state levels decrease. These results further indicate that the binding of $PKM\zeta$ and KIBRA, which results in the formation of droplets, inhibits the elimination of these proteins from the cell. On the basis of these observations we developed a more detailed phenomenological negative-feedback model to account for maintenance.

The phenomenological model (Fig. 5) includes 4 distinct species. These are $PKM\zeta$, K- which stands for KIBRA, X which denotes dimers of $PKM\zeta$ bound to K, and Y, which represents aggregates, clusters, or droplets of X hetero-dimers. The exact identity of Y is not determined here, and therefore this model is phenomenological. The species Y could represent large complexes of bound X via KIBRA-KiBRA interactions, or even droplets or aggregates of X. However, we must clarify that a 3 species model without a species such as Y, and in which elementary kinetics determine the formation of X would not be bi-stable, and therefore would not account for maintenance. All kinetics are standard mass-action



$$PKM\zeta + K \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} X \quad (R1)$$

$$X + X \underset{k_{-2}}{\overset{F}{\rightleftarrows}} Y \quad (R2)$$

$$X + Y \underset{k_{-3}}{\overset{F}{\rightleftarrows}} Y \quad (R3)$$

$$0 \underset{\lambda_M}{\overset{s_M}{\rightleftarrows}} PKM\zeta \quad (R4)$$

$$0 \underset{\lambda_K}{\overset{s_K}{\rightleftarrows}} K \quad (R5)$$

$$X \underset{\lambda_X}{\rightarrow} 0 \quad (R6)$$

$$Y \underset{\lambda_Y}{\rightarrow} 0 \quad (R7)$$

Figure 5: Set-up of phenomenological model. Left - a schematic diagram of the model. Arrows on the top represent the source term due to protein synthesis of both $PKM\zeta$ and KIBRA. Dashed lines with arrows at the bottom represent protein elimination, which is slower for the clusters or aggregates (Y). Right - detailed description of all the reaction equations, representing the reaction diagram on the left.

elementary kinetics (as described in appendix A equations 8- 11), except for the formation of the species Y, which depends on a non-linear Hill-function F (Fig. 5, see equation 12 in appendix A for formal definition). This steep function implements (but does not explain) the ultrasensitive reaction necessary for the a negative-feedback model. The simple binding of $PKM\zeta$ to K uses mass-action kinetics, which are not ultrasensitive. Formation of Y could require cooperativity, or possibly a phase transition from a diffuse phase to a liquid-liquid separated phase. While the experimental data provides some support for such a process, currently there is insufficient evidence for modeling it explicitly. However, for the purposes of this paper it is not essential to understand the origin of this non-linearity. A critical component of this model is that the degradation of Y is much slower than of the other species. It is this assumption that puts this model into the negative-feedback framework, and is necessary for bi-stability. The reactions arising from this setup are depicted in Figure 5, where $F = F(X, Y)$ is the non-linear ultrasensitive function; an example of which is given in the appendix.

In Figure 6A we show results of simulations in which L-LTP is induced via a transient increase in the source term of $PKM\zeta$ and KIBRA. These simulations demonstrate that this phenomenological negative-feedback model is indeed bi-stable, and that an input pulse of $PKM\zeta$ and KIBRA is sufficient to shift it from the DOWN state to the UP state. This input pulse is equivalent to a transient increase in protein synthesis of key synaptic proteins which we know is essential for the induction of L-LTP in genera l(Klann and Sweatt, 2008). The allosteric inhibitors $\zeta - stat$ or K-ZAP that act by inhibiting the binding of $PKM\zeta$ to $KIBRA$ reverses L-LTP and long-term memory (Tsokas et al., 2024). We simulated the application of ζ-stat or K-ZAP by transiently reducing the binding coefficient between $PKM\zeta$ and $KIBRA$. We show in Figure 6A that this is sufficient for shifting the system from the UP state back to the DOWN state. If the input pulse is smaller (Fig. 6B) L-LTP is not induced, hence there is a threshold for the induction of L-LTP in the model. We also show (Fig. 6B) that the application of ζ-stat or K-ZAP at the basal level causes only a minor and transient change in protein levels, this is also consistent with experimental results (Tsokas et al., 2024).

The ability to obtain bi-stability in this model critically depends on the ultrasensitive function F. In the simulation of Figure 6 we use a Hill-function with a coefficient $n = 4$ (defined in equation 12). For the same set of parameters we find that we cannot establish bi-stability for $n < 2.5$. This result is based on simulations, but since we have not fully analyzed the system it is possible that one could obtain bi-stability for smaller values of n for other parameters sets. However, independent of the exact quantitative details, the qualitative properties of the system are reminiscent to those of the simple model, and similarly it critically depends on an ultrasensitive reaction. A full understanding of such a system would also require identifying and understanding the mechanistic origin of this necessary ultrasensitivity.

Previously we have shown analytically for the simple model with a concentration dependent degradation rate (equation 3), that the ratio of protein concentrations in the UP

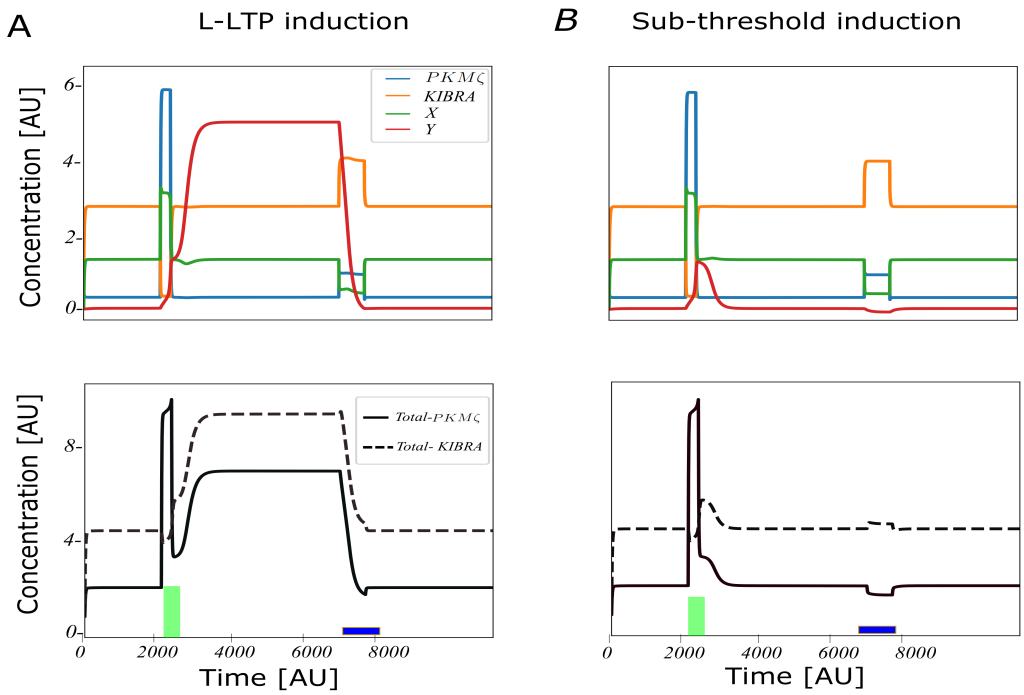


Figure 6: Dynamics of phenomenological negative-feedback model. A. Dynamics of L-LTP induction and reversal. Top - all species, Bottom - total $PKM\zeta$ and $KIBRA$. Initially the model converges to a sustained basal level sate (DOWN-state). An induction stimulus (green bar) causes a large and sustained increase in Y, and as a consequence also in total $PKM\zeta$ and $KIBRA$ (UP-state).cShown experimentally in Tsokas et al., (2024). Simulated ζ-stat or K-ZAP (blue bar) causes a reversal to the basal level. B. A weaker induction stimulus causes only a transient increase of $PKM\zeta$ and X, but no increase in Y. This sub-threshold stimulus fails to induce L-LTP. Similarly application of $\zeta - stat$ at baseline causes only a minor transient change. These are deterministic simulations with zero noise added.

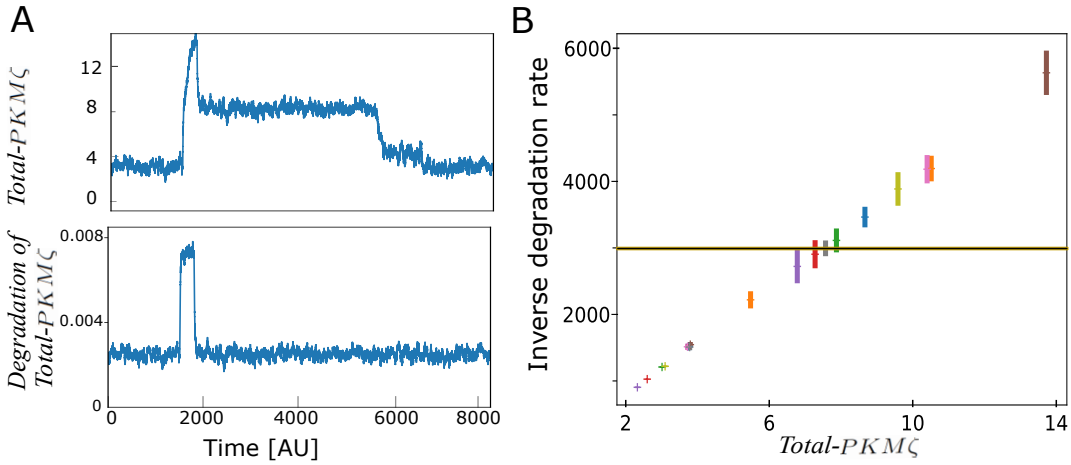


Figure 7: Inverse degradation as function of protein levels in the phenomenological model. **A.** Dynamics of L-LTP induction and reversal with added noise. Top- Total $PKM\zeta$; Bottom - degradation of total-$PKM\zeta$. **B.** Inverse degradation rate of $PKM\zeta$ vs. steady state levels of total $PKM\zeta$ in UP and DOWN states for different parameters. Each color represents different parameter sets, different data points with same color are from the same parameters but for the UP and DOWN states. All different parameter sets shown have the same synthesis rate. The error bars are standard deviations due to noise added to the system. The solid black line is what would be expected from a model that depends purely on positive feedback, and the Y intercept of this line is $1/\lambda$

and DOWN states are inversely proportional to the ratio of their turnover rates. In some sense this must be true in general, since it is clearly the balance between production and degradation that sets steady-state levels, and since in this model the production level is constant, then the only possible source of change determining steady state levels is the effective degradation rate. Nevertheless it is instructive to computationally measure this for the simulations of the model as a benchmark that can be compared to experiments to test if they indeed fall into the negative-feedback framework. To do this we calculated the a degradation rate coefficient of $PKM\zeta$ in the UP and DOWN states, The degradation rate coefficient was calculated by measuring the total turnover of $PKM\zeta$ from each one of the species that include molecules of $PKM\zeta$ and divided this by the total concentration of $PKM\zeta$ in each of those states (see appendix A). In the abstract model of section **??** this calculation would produce a the value of λ_1 in the DOWN state and λ_2 in the UP state.

In Figure 7A we show the total level of $PKM\zeta$ while we shift the system from the DOWN to UP state and back again (top). We also show the total degradation level $PKM\zeta$ in those during the same period (bottom). While the level of total $PKM\zeta$ increases from the DOWN to UP states, the degradation rate is similar in both, only transiently increasing during the induction phase. We added noise to these simulations to better

emulate physiological conditions, this noise adds fluctuations to the simulations shown in Figure 6A, but does not affect bi-stability for this level of noise and for the parameters used here. We repeated these simulations for various parameter sets in which bi-stability is exhibited. In Figure 7B we plot the inverse of the degradation rate as a function of the total protein for the $PKM\zeta$ species for these different parameters. As can be seen, the expected linear relationship holds. We show the inverse degradation rather than the degradation rate itself because it results in simpler linear-curves. For all parameter sets presented in this figure the same synthesis rates were used. Using different synthesis rates would result in a different linear curve for each synthesis rate, with a different slope for each synthesis rate. Because we are showing inverse degradation, higher synthesis rates produce smaller slopes and lower synthesis rate produce larger slopes (results not shown). In contrast to this linear relationship exhibited in the negative-feedback model, in a positive feedback model the inverse degradation coefficient is independent of state of the system. In the case of the simple positive feedback of section **??** the inverse degradation rate is simply $1/\lambda$ as represented by the solid black line of Figure 7B. Note that all data points in Figure 7B are from a model with different parameters, but the same production rates for both $PKM\zeta$ and KIBRA. With a different production rates the behavior would be qualitatively similar, but the slope would be different.

The relationship between synaptic efficacy and degradation could be measured experimentally in order to test which of the different theories is consistent with the data. Specifically, the negative feedback model would predict that after a synapse undergoes L-LTP, the degradation rate of key proteins such as $PKM\zeta$ or KIBRA would decrease. In a positive-feedback model it would not change. However, these different mechanisms are not necessarily mutually exclusive, and it is possible that both mechanisms are used in the same synapses, i.e., that stronger synapses might exhibit both less degradation and more synthesis. Experimental data could indicate if maintenance is accomplished via a pure positive-feedback or negative-feedback model, or via a hybrid model. If a hybrid model is used, the maintenance system might exhibit multi-stability; enhancing the dynamic range of a synapse.

3 Discussion

In this paper we propose a new class of models to account for the stability of synaptic efficacies. We show that models based on negative feedback of protein elimination could generate synaptic stability, and that these result in predictions that can be tested experimentally. This class of models are based on two key assumptions. First, in a potentiated state key synaptic proteins are eliminated at a slower rate than in the un-potentiated state. Second, there is a steep hypersensitive transition separating the un-potentiated and potentiated states. Such models account for experimental observations, and generate predictions that can be experimentally tested.

The premise that stable synaptic efficacies underlie stable memories and learning has been called the synaptic trace theory (Mongillo et al., 2017). However, recent long-term recording of synapses in vivo using two-photon microscopy has found a large scale turnover of synaptic spines and large changes in their size over a period of several weeks in the CA1 subregion in the hippocampus (Attardo et al., 2015). Experiments that show synaptic volatility seem contradict the assumption of the synaptic trace theory. In contrast, spines seem to be much more stable in adult sensory and motor cortex (Grutzendler et al., 2002; Zuo et al., 2005). Despite the recent advances in recording techniques we do not yet know if the stable synapses observed are indeed those crucial for the maintenance of memory.

The relative apparent instability of synaptic efficacies is mirrored by instability of neuronal representations in some brain systems and conditions (Ziv et al., 2013) but not others (Refaeli et al., 2023; Pérez-Ortega et al., 2021). A drift in neuronal representations might be a consequence of synaptic instability. Models have shown that in some systems a drift of the input representation can be compensated for upstream, and slowed down by external feedback (Rule et al., 2020) or by internal mechanisms (Pérez-Ortega et al., 2021). Such mechanisms require constant re-exposure to the same environment, and this does not apply to memory of old events that one is not re-exposed to. Currently it is unclear how old memories can be retained if there is no stable synaptic core. Many synapses might still drift, if their drift is in a subspace orthogonal to the memory, but currently it seems likely an episodic memory cannot be maintained without a stable synaptic core (Clopath et al., 2017).

Another possible objection is that a process of system-level consolidation occurs after the induction of long-term memory. In system consolidation the location of memory storage is shifted from the hippocampal region to cortical regions (Wang and Morris, 2010; Squire et al., 2015). However, in these cortical regions memory is also stored via synaptic plasticity, which means system level consolidation does not change the nature of the problem, but only its physical location. Despite system level consolidation, it has been shown that engrams in CA1 and neocortex are highly stable for long period of time, consistent with memory storage (Refaeli et al., 2023; Lee et al., 2023). More directly, we have specifically shown that the reversal of L-LTP in the hippocampus can disrupt memory storage up to at least one month after the induction of memory (Tsokas et al., 2024). Showing that at least some memories are stored in the hippocampus for at least a month (Tsokas et al., 2024), much longer than typical dwell times of key synaptic proteins.

One must note that models of positive feedback are the prevalent models not only for memory maintenance, but more prominently for models of cell fate (Kobayashi et al., 2003; Zhu et al., 2022). In those systems the feedback loop is typically implemented at the level of transcription. In these theories transcribed proteins either directly or indirectly modulate their own transcription. Models implemented at the level of transcription are whole cell models. Therefore, they are inappropriate for maintenance of synaptic plasticity that must be synapse specific in order to be able to maintain specific memories, generate selective receptive fields, and for specific learning in general. Although, a transcription

based feedback loop is inappropriate as the primary mechanism for maintenance of synaptic plasticity, changes in transcription may also occur during maintenance (Squire et al., 2015).

In order to serve as a basis for memory formation, and for learning in general, synaptic plasticity must exhibit a high degree of synaptic specificity. Synaptic specificity means that while some synapses are potentiated, neighboring synapses remain un-potentiated. Such synaptic specificity has been observed experimentally during the induction of LTP (Harvey and Svoboda, 2007) and L-LTP (Govindarajan et al., 2011). Positive-feedback models of maintenance are based on mechanisms that are in principle local such as post-translational modifications, or local translation of proteins. However, until recently the degree of synapse specificity arising from such supposedly local mechanisms has not been analyzed theoretically. To address this question we recently developed a reaction diffusion model of maintenance based on positive feedback (Huertas et al., 2022). The model includes a dendrite with many synaptic spines and with molecular switches based on positive-feedback. We studied this model using both analytical and numerical methods, and we have shown that in order to obtain synaptic specificity it is necessary that the switches reside in synaptic spines. However, although necessary this might not be sufficient, as we have also shown that even if the bi-stable switches reside in spines it is very difficult to obtain synaptic specificity at the level observed experimentally. This lack of specificity occurs because proteins generated in active spines can spill over to neighboring inactive spines and turn on the switches in those spines. If a sufficient number of such synapses are in the potentiated state this spillover will cause all their neighbors within the dendritic branch to potentiate as well, resulting in a loss synaptic specificity. Our results showed that the one way to obtain a realistic level of synaptic selectivity is if in addition to the positive feedback-loop another mechanism during the induction of L-LTP restricts the diffusion of the critical protein out of the potentiated spine. Structural changes that occur in L-LTP (Fukazawa et al., 2003) suggest that such changes in diffusion are possible, but we have shown that such a change needs to be quantitively quite large in order to establish realistic levels of synaptic specificity. It is reasonable to expect that a negative-feedback model will not suffer from the same limits to synaptic specificity. This is because in such a model potentiated synapses do not persistently generate additional proteins, which might then spill over. However, we have not yet analyzed this theoretically.

The theoretical framework proposed here is based on two key assumption, the local control of protein elimination, and a mechanism that allows a sharp transition between the two states; the fast and slow protein elimination states. The key question of what these mechanisms are has not been solved here. However, the observation that maintenance depends on the binding of $PKM\zeta$ and KIBRA suggests elements of these mechanisms. First, it has already been shown that the binding of KIBRA to $PKM\zeta$ slows the proteasomal degradation of $PKM\zeta$. This result points in the right direction but might not be sufficient because diffusion rather than protein turnover might be the limiting factor. However, results in an expression system (Supplementary Figure) show that bound $PKM\zeta$ and KIBRA segregate into droplets (Supplementary Figure A).

The existence of these droplets depends on the binding as it does not exist if each protein is expressed independently, and is eliminated by inhibitors of this binding. These droplets resemble liquid-liquid phase separation within synaptic spine (Liu et al., 2021; Chen et al., 2020). If these droplets are photo-bleached, their fluorescence recovers (results not shown), indicating they are indeed in a liquid-like state. Such larger droplets are likely to diffuse at a much more slowly than independent proteins, especially in the crowded synaptic compartment, further contributing to a slower elimination rate. The formation of these droplets might depend on a highly non-linear cooperative process such as the one proposed in section 2.4. Such processes might depend on interactions between heterodimers through KIBRA-KIBRA interactions. Another way of looking at this is that there is a possible phase transition within potentiated synapses from a diffuse state in which these droplets do not form, to a phase separated state. If there is indeed a concentration dependent phase transition, during the formation of long-term potentiated synapses, this could be the mechanism for the required ultrasensitive component. Although our results suggest a phase separation, we have not yet demonstrated a phase transition, either in the expression system or in live synaptic spines.

APPENDIX

A Equations of Phenomenological model

The complete set of equations have the form:

$$\frac{dPKM\zeta}{dt} = -k_1 PKM\zeta \cdot K + k_{-1}X - \lambda_{PKM\zeta} \cdot PKM\zeta + I_{PKM\zeta} \quad (8)$$

$$\frac{dK}{dt} = -k_1 PKM\zeta \cdot K + k_{-1}X - \lambda_K \cdot K + I_K(1 + n_K \cdot \eta(t)) \quad (9)$$

$$\frac{dX}{dt} = -k_{-1}X + k_1 PKM\zeta \cdot K - \rho F(X, Y, par) + k_{-Y}Y - \lambda_X X \quad (10)$$

$$\frac{dY}{dt} = -k_{-Y}Y + \rho F(X, Y, par) - \lambda_Y Y \quad (11)$$

where:

$$F(X, Y, \overline{par}) = c_1 \frac{X^2}{K_{XX}^2 + X^2} + c_2 \frac{(X \cdot Y)^n}{K_{XY}^n + (X \cdot Y)^n}, \quad (12)$$

in which $\overline{par}$ represents the parameters: $c_1, c_2, K_{XX}, K_{XY}, n$. The non-linear saturating equation defining F is not derived from elementary kinetics, and should be seen as a purely phenomenological term, in that it enables the system of equations to produce the required phenomenology. We have not, and are not attempting here to account for how the biochemistry or physics of this system might generate this term.

In the simulations presented here these parameters are: $\lambda_X = 0.1, \lambda_Y = 0.00001, \lambda_K = 0.075, \lambda_{PKM\zeta} = 0.15$, $k_1 = 0.25, k_{-1} = 0.1, k_{-Y} = 0.01$, $n = 4$, $c_1 = 0.05, c_2 = 0.25$, $K_{XX} = 2.5, K_{XY} = 4$. The currents $I_{PKM\zeta}$ and I_K include a steady state components, and a transient component for the induction of L-LTP. Here we assume that only $I_{PKM\zeta}$ has a transient component. The steady state components used here are: $I_{PKM\zeta} = 0.35, I_K = 0.2$. The transient component for inducing L-LTP was 1.0 for a duration of 200 arbitrary units of time. The application of ζ-stat or K-ZAP is emulated by significantly reducing k_1 transiently. Here we set $k_1 = 0.025$ for limited time window; a tenth of its basal value. Noise is added phenomenologically using Langevin-like noise. The term η generates uncorrelated Gaussian noise, with mean zero and variance 1, and $n_{PKM\zeta}, n_k$ and noise parameters for $PKM\zeta$ and K, respectively. There is a large range of parameter that has qualitatively the same behavior. The code for running these simulation, is written in Julia and will be uploaded on ModelDB after the paper is accepted for publication.

To calculate the degradation coefficient of $PKM\zeta$, we first calculated the total degradation of $PKM\zeta$ at each time step (dt), and for this model it is: $dt(\lambda_{PKM\zeta} \cdot PKM\zeta + \lambda_X \cdot X + \lambda_Y \cdot Y)$. These are the degradation levels used in as shown in Figure 7A (bottom). To obtain the degradation rate coefficient we divided the total degradation of $PKM\zeta$ at each

time point by the total concentration of $PKM\zeta$ (Fig. 7A, top) at the same time point, and averaged separately over the time points in the DOWN and UP states, while omitting the transients. The y axis values of Figure 7B are 1 over this degradation rate coefficient, and the error bars are standard deviation of the inverse coefficient over those time points.

References

Agarwal A, Adams R, Castellani GC, Shouval HZ (2012) On the precision of quasi steady state assumptions in stochastic dynamics. *The Journal of Chemical Physics* 137:044105.

Aslam N, Kubota Y, Wells D, Shouval HZ (2009) Translational switch for long-term maintenance of synaptic plasticity. *Molecular Systems Biology* 5:284 Publisher: John Wiley & Sons, Ltd.

Aslam N, Shouval HZ (2012) Regulation of cytoplasmic polyadenylation can generate a bistable switch. *BMC Systems Biology* 6:12.

Attardo A, Fitzgerald JE, Schnitzer MJ (2015) Impermanence of dendritic spines in live adult CA1 hippocampus. *Nature* 523:592–596 Publisher: Nature Publishing Group.

Bear MF, Malenka RC (1994) Synaptic plasticity: LTP and LTD. *Current Opinion in Neurobiology* 4:389–399.

Bhalla US, Iyengar R (1999) Emergent properties of networks of biological signaling pathways. *Science* 283:381–387 Publisher: American Association for the Advancement of Science.

Chen X, Wu X, Wu H, Zhang M (2020) Phase separation at the synapse. *Nature neuroscience* 23:301–310.

Clopath C, Bonhoeffer T, Hübener M, Rose T (2017) Variance and invariance of neuronal long-term representations. *Philosophical Transactions of the Royal Society B: Biological Sciences* 372:20160161.

Crick F (1984) Neurobiology: Memory and molecular turnover. *Nature* 312:101–101 Publisher: Nature Publishing Group.

Fukazawa Y, Saitoh Y, Ozawa F, Ohta Y, Mizuno K, Inokuchi K (2003) Hippocampal LTP Is Accompanied by Enhanced F-Actin Content within the Dendritic Spine that Is Essential for Late LTP Maintenance In Vivo. *Neuron* 38:447–460.

Gabbiani F, Cox SJ (2017) *Mathematics for neuroscientists* Academic Press.

Gillespie DT (2002) Exact stochastic simulation of coupled chemical reactions Archive Location: world Publisher: American Chemical Society.

Govindarajan A, Israely I, Huang SY, Tonegawa S (2011) The dendritic branch is the preferred integrative unit for protein synthesis-dependent LTP. *Neuron* 69:132–146 Publisher: Elsevier.

Grutzendler J, Kasthuri N, Gan WB (2002) Long-term dendritic spine stability in the adult cortex. *Nature* 420:812–816 Publisher: Nature Publishing Group.

Harvey CD, Svoboda K (2007) Locally dynamic synaptic learning rules in pyramidal neuron dendrites. *Nature* 450:1195–1200 Publisher: Nature Publishing Group.

Hsieh C, Tsokas P, Grau-Perales A, Lesburguères E, Bukai J, Khanna K, Chorny J, Chung A, Jou C, Burghardt NS et al. (2021) Persistent increases of pkmζ in memory-activated neurons trace ltp maintenance during spatial long-term memory storage. *European Journal of Neuroscience* 54:6795–6814.

Huertas MA, Newton AJH, McDougal RA, Sacktor TC, Shouval HZ (2022) Conditions for Synaptic Specificity during the Maintenance Phase of Synaptic Plasticity. *eNeuro* 9:ENEURO.0064–22.2022.

Jalil SJ, Sacktor TC, Shouval HZ (2015) Atypical PKCs in memory maintenance: the roles of feedback and redundancy. *Learning & memory* 22:344–353 Publisher: Cold Spring Harbor Lab.

Klann E, Sweatt JD (2008) Altered protein synthesis is a trigger for long-term memory formation. *Neurobiology of Learning and Memory* 89:247–259.

Kobayashi T, Chen L, Aihara K (2003) Modeling Genetic Switches with Positive Feedback Loops. *Journal of Theoretical Biology* 221:379–399.

Langille JJ, Brown RE (2018) The Synaptic Theory of Memory: A Historical Survey and Reconciliation of Recent Opposition. *Frontiers in Systems Neuroscience* 12.

Lee JH, Kim WB, Park EH, Cho JH (2023) Neocortical synaptic engrams for remote contextual memories. *Nature Neuroscience* 26:259–273 Publisher: Nature Publishing Group.

Ling DSF, Benardo LS, Serrano PA, Blace N, Kelly MT, Crary JF, Sacktor TC (2002) Protein kinase M ζ is necessary and sufficient for LTP maintenance. *Nature Neuroscience* 5:295–296 Bandiera_abtest: a Cg_type: Nature Research Journals Number: 4 Primary_atype: Research Publisher: Nature Publishing Group.

Lisman JE (1985) A mechanism for memory storage insensitive to molecular turnover: a bistable autophosphorylating kinase. *Proceedings of the National Academy of Sciences* 82:3055–3057 Publisher: National Acad Sciences.

Lisman JE, Zhabotinsky AM (2001) A Model of Synaptic Memory: A CaMKII/PP1 Switch that Potentiates Transmission by Organizing an AMPA Receptor Anchoring Assembly. *Neuron* 31:191–201.

Liu PW, Hosokawa T, Hayashi Y (2021) Regulation of synaptic nanodomain by liquid–liquid phase separation: A novel mechanism of synaptic plasticity. *Current Opinion in Neurobiology* 69:84–92.

Martin SJ, Grimwood PD, Morris RG (2000) Synaptic plasticity and memory: an evaluation of the hypothesis. *Annual review of neuroscience* 23:649–711 Publisher: Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA.

Miller P, Zhabotinsky AM, Lisman JE, Wang XJ (2005) The Stability of a Stochastic CaMKII Switch: Dependence on the Number of Enzyme Molecules and Protein Turnover. *PLOS Biology* 3:e107 Publisher: Public Library of Science.

Mongillo G, Rumpel S, Loewenstein Y (2017) Intrinsic volatility of synaptic connections—a challenge to the synaptic trace theory of memory. *Current opinion in neurobiology* 46:7–13 Publisher: Elsevier.

Nicoll RA (2017) A brief history of long-term potentiation. *neuron* 93:281–290 Publisher: Elsevier.

Ogasawara H, Kawato M (2010) The protein kinase Mζ network as a bistable switch to store neuronal memory. *BMC Systems Biology* 4:1–10 Publisher: Springer.

Pastalkova E, Serrano P, Pinkhasova D, Wallace E, Fenton AA, Sacktor TC (2006) Storage of Spatial Information by the Maintenance Mechanism of LTP. *Science* 313:1141–1144 Publisher: American Association for the Advancement of Science.

Pérez-Ortega J, Alejandre-García T, Yuste R (2021) Long-term stability of cortical ensembles. *eLife* 10:e64449 Publisher: eLife Sciences Publications, Ltd.

Refaeli R, Kreisel T, Groysman M, Adamsky A, Goshen I (2023) Engram stability and maturation during systems consolidation. *Current Biology* 33:3942–3950.e3 Publisher: Elsevier.

Richter JD, Klann E (2009) Making synaptic plasticity and memory last: mechanisms of translational regulation. *Genes & development* 23:1–11 Publisher: Cold Spring Harbor Lab.

Rule ME, Loback AR, Raman DV, Driscoll LN, Harvey CD, O'Leary T (2020) Stable task information from an unstable neural population. *elife* 9:e51121 Publisher: eLife Sciences Publications, Ltd.

Sacktor TC, Osten P, Valsamis H, Jiang X, Naik MU, Sublette E (1993) Persistent activation of the zeta isoform of protein kinase C in the maintenance of long-term potentiation. *Proceedings of the National Academy of Sciences* 90:8342–8346.

Squire LR, Genzel L, Wixted JT, Morris RG (2015) Memory Consolidation. *Cold Spring Harbor Perspectives in Biology* 7:a021766 Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

Tsokas P, Hsieh C, Flores-Obando RE, Bernabo M, Tcherepanov A, Hernández AI, Thomas C, Bergold PJ, Cottrell JE, Kremerskothen J, Shouval HZ, Nader K, Fenton AA, Sacktor TC (2024) KIBRA anchoring the action of PKM ζ maintains the persistence of memory. *Science Advances* 10:eadl0030.

Vogt-Eisele A, Krüger C, Duning K, Weber D, Spoelgen R, Pitzer C, Plaas C, Eisenhardt G, Meyer A, Vogt G, Krieger M, Handwerker E, Wennmann DO, Weide T, Skryabin BV, Klugmann M, Pavenstädt H, Huentelmann MJ, Kremerskothen J, Schneider A (2014) KIBRA (KIdney/BRAin protein) regulates learning and memory and stabilizes Protein kinase Mζ. *Journal of Neurochemistry* 128:686–700 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jnc.12480.

Wang SH, Morris RG (2010) Hippocampal-Neocortical Interactions in Memory Formation, Consolidation, and Reconsolidation. *Annual Review of Psychology* 61:49–79.

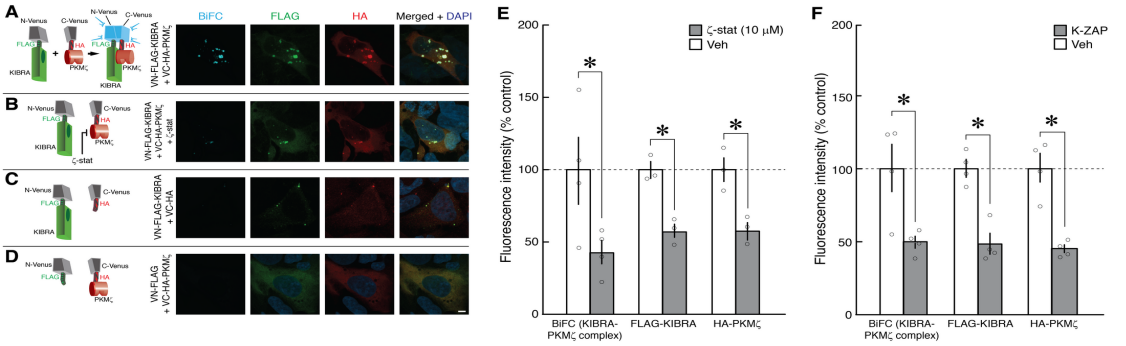
Whitlock JR, Heynen AJ, Shuler MG, Bear MF (2006) Learning induces long-term potentiation in the hippocampus. *science* 313:1093–1097 Publisher: American Association for the Advancement of Science.

Yang Y, Liu Dq, Huang W, Deng J, Sun Y, Zuo Y, Poo Mm (2016) Selective synaptic remodeling of amygdalocortical connections associated with fear memory. *Nature Neuroscience* 19:1348–1355 Number: 10 Publisher: Nature Publishing Group.

Zhu R, Del Rio-Salgado JM, Garcia-Ojalvo J, Elowitz MB (2022) Synthetic multistability in mammalian cells. *Science (New York, N.Y.)* 375:eabg9765.

Ziv Y, Burns LD, Cocker ED, Hamel EO, Ghosh KK, Kitch LJ, Gamal AE, Schnitzer MJ (2013) Long-term dynamics of CA1 hippocampal place codes. *Nature Neuroscience* 16:264–266 Publisher: Nature Publishing Group.

Zuo Y, Lin A, Chang P, Gan WB (2005) Development of Long-Term Dendritic Spine Stability in Diverse Regions of Cerebral Cortex. *Neuron* 46:181–189 Publisher: Elsevier.



Supplementary Figure 1: KIBRA-$PKM\zeta$ antagonists decrease decrease KIBRA-$PKM\zeta$ interactions and lower total KIBRA and $PKM\zeta$ levels. (A-D) Left, schematics of Venus-fusion constructs used in BiFC experiments. Right, representative images of HEK293T cells 1 day after transfection. Left column, BiFC; middle columns, immunocytochemistry for FLAG-tagged KIBRA and HA-tagged $PKM\zeta$; right column, merged images with nuclear stain DAPI, in which for clarity the BiFC signal is blue. (A) BiFC produced by co-transfection of N-terminal-Venus FLAG-tagged KIBRA and C-terminal-Venus HA-tagged $PKM\zeta$ shows accumulation of intracellular droplets of KIBRA-$PKM\zeta$ complexes. (B ζ-stat (10 µM) decreases BIFC, and total levels of KIBRA and $PKM\zeta$. (C) Co-transfection of the N-terminal-Venus FLAG-tagged KIBRA and the C-terminal-Venus tagged with HA but without $PKM\zeta$ shows minimal BiFC. Transfection of KIBRA alone shows fewer puncta and less immunofluorescence than KIBRA together with $PKM\zeta$. (D) Co-transfection of C-terminal-Venus HA-tagged $PKM\zeta$ and N-terminal-Venus tagged with FLAG but without KIBRA shows minimal BiFC. Transfection of $PKM\zeta$ alone shows diffuse cytosolic distribution and no detectable puncta without KIBRA. Scale bar, 5 µm. Right, mean ± SEM shows ζ-stat (E) and K-ZAP (F) decrease BiFC, total KIBRA, and total $PKM\zeta$ immunofluorescence intensities (E: t3 = 3.39, P = 0.04, d = 1.69; t2 = 10.09, P = 0.010, d = 5.82; t2 = 12.69, P = 0.006, d = 7.33; F: t3 = 4.00, P = 0.03, d = 2.00; t3 = 6.24, P = 0.008, d = 3.12; t3 = 6.08, P = 0.009, d = 3.04, respectively). * denotes P ¡ 0.05. Experiments described in Tsokas et al., 2024. interactions and lower total KIBRA and $PKM\zeta$ levels. (A-D) Left, schematics of Venus-fusion constructs used in BiFC experiments. Right, representative images of HEK293T cells 1 day after transfection. Left column, BiFC; middle columns, immunocytochemistry for FLAG-tagged KIBRA and HA-tagged $PKM\zeta$; right column, merged images with nuclear stain DAPI, in which for clarity the BiFC signal is blue. (A) BiFC produced by co-transfection of N-terminal-Venus FLAG-tagged KIBRA and C-terminal-Venus HA-tagged $PKM\zeta$ shows accumulation of intracellular droplets of KIBRA-$PKM\zeta$ complexes. (B)ζ-stat (10 µM) decreases BIFC, and total levels of KIBRA and $PKM\zeta$. (C) Co-transfection of the N-terminal-Venus FLAG-tagged KIBRA and the C-terminal-Venus tagged with HA but without $PKM\zeta$ shows minimal BiFC. Transfection of KIBRA alone shows fewer puncta and less immunofluorescence than KIBRA together with $PKM\zeta$. (D) Co-transfection of C-terminal-Venus HA-tagged $PKM\zeta$ and N-terminal-Venus tagged with FLAG but without KIBRA shows minimal BiFC. Transfection of $PKM\zeta$ alone shows diffuse cytosolic distribution and no detectable puncta without KIBRA. Scale bar, 5 μm. Right, mean ± SEM shows ζ-stat (E) and K-ZAP (F) decrease BiFC, total KIBRA, and total $PKM\zeta$ immunofluorescence intensities (E: t3 = 3.39, P = 0.04, d = 1.69; t2 = 10.09, P = 0.010, d = 5.82; t2 = 12.69, P = 0.006, d = 7.33; F: t3 = 4.00, P = 0.03, d = 2.00; t3 = 6.24, P = 0.008, d = 3.12; t3 = 6.08, P = 0.009, d = 3.04, respectively). * denotes P < 0.05. Experiments described in Tsokas et al., 2024.