The Binomial Model in Fluctuation Analysis of Quantal Neurotransmitter Release

D. M. J. Quastel

Department of Pharmacology and Therapeutics, Faculty of Medicine, The University of British Columbia, Vancouver, British Columbia V6T 1Z3 Canada

ABSTRACT The mathematics of the binomial model for quantal neurotransmitter release is considered in general terms, to explore what information might be extractable from statistical aspects of data. For an array of N statistically independent release sites, each with a release probability p, the compound binomial always pertains, with $\langle m\rangle = N\langle p\rangle$, $p' \equiv 1 - \mathrm{var}(m)/\langle m\rangle = \langle p\rangle(1 + cv_p^2)$ and $n' \equiv \langle m\rangle/p' = N/(1 + cv_p^2)$, where m is the output/stimulus and cv_p^2 is $\mathrm{var}(p)/\langle p\rangle^2$. Unless n' is invariant with ambient conditions or stimulation paradigms, the simple binomial ($cv_p = 0$) is untenable and n' is neither N nor the number of "active" sites or sites with a quantum available. At each site $p = p_o p_A$, where p_o is the output probability if a site is "eligible" or "filled" despite previous quantal discharge, and p_A (eligibility probability) depends at least on the replenishment rate, p_o, and interstimulus time. Assuming stochastic replenishment, a simple algorithm allows calculation of the full statistical composition of outputs for any hypothetical combinations of p_o's and refill rates, for any stimulation paradigm and spontaneous release. A rise in n' (reduced cv_p) tends to occur whenever p_o varies widely between sites, with a raised stimulation frequency or factors tending to increase p_o's. Unlike $\langle m\rangle$ and $\mathrm{var}(m)$ at equilibrium, output changes early in trains of stimuli, and covariances, potentially provide information about whether changes in $\langle m\rangle$ reflect change in $\langle p_o\rangle$ or in $\langle p_A\rangle$. Formulae are derived for variance and third moments of postsynaptic responses, which depend on the quantal mix in the signals. A new, easily computed function, the area product, gives noise-unbiased variance of a series of synaptic signals and its peristimulus time distribution, which is modified by the unit channel composition of quantal responses and if the signals reflect mixed responses from synapses with different quantal time course.

INTRODUCTION

From the first description of the quantal nature of neurotransmitter release and its probabilistic character it has been generally assumed that the distribution of numbers of quanta released by stimuli is in some sense binomial in character, with a Poisson distribution appearing under conditions of depressed release (low Ca^{2+}/raised Mg^{2+}), so that each stimulus effectively samples with low probability from a relatively large number of release sites (del Castillo and Katz, 1954a; Martin, 1955). Since then many studies of a variety of synapses have found deviations from a Poisson distribution in the direction expected with a binomial distribution (rev. McLachlan, 1978; Redman, 1990).

A binomial distribution for outputs corresponds to the a priori consideration that the number of quanta released by a stimulus must be limited by the number of release sites and the number of quanta available for release. Equally a priori, however, there is no reason to assume that the probability of release is the same for every site and that this probability is the same from one stimulus to the next, which are both preconditions for a simple binomial distribution of outputs for which the mean ($\langle m\rangle$) is np and the variance, $\mathrm{var}(m)$, is $np(1 - p)$, where n is the number sampled with probability p.

The case in which p and n may vary spatially and temporally was considered by Brown et al. (1976), who showed that the p and n obtained from data assuming a simple binomial (i.e., $p' = 1 - \mathrm{var}(m)/\langle m\rangle$ and $n' = \langle m\rangle/p'$) then bear no relation to the true mean p ($\langle p\rangle$) and the true (mean) number of sites capable of release, except that n', like the number of "filled" or "eligible" sites, must be less than the total number of sites, and p', like any p, must be ≤ 1. Given that at each release site p must be the product of output probability (p_o) and the probability that the site is "eligible" (p_A), the simple binomial (all release sites equivalent) should give n' always equal to the total number of release sites. Numerous findings that n' varies with stimulation frequency and ambient Ca^{2+}/Mg^{2+} (see McLachlan, 1978) therefore indicate that the simple binomial does not apply generally. Nevertheless, it is still common to find in the literature the assumption that n' somehow represents the number of release sites that have a quantum available for release.

It is my purpose here to consider the binomial model with minimal artificial constraints, emphasizing the underlying assumptions and the way in which these translate mathematically into the consequent statistical properties of outputs and resulting postsynaptic signals. A synapse or group of synapses is (conventionally) envisaged as an array of N independent release sites, each potentially capable of releasing only one quantum at a time, and each with its own p. Unless all release sites have the same p, giving the simple binomial model, this is the compound binomial model (rev. McLachlan, 1978; rev. Redman, 1990; Dityatev, Kozhanov and Gapanovich, 1992), for which $\langle m\rangle$ is $N\langle p\rangle$, $p' = \langle p\rangle$

Received for publication 25 September 1995 and in final form 1 October 1996.

Address reprint requests to Dr. D. M. J. Quastel, Department of Pharmacology and Therapeutics, Faculty of Medicine, The University of British Columbia, 2176 Health Sciences Mall, Vancouver BC V6T 1Z3, Canada. Tel.: 604-822-4263; Fax: 604-822-6012; E-mail: puil@neuro.pharmacology.ubc.ca.

0006-3495/97/02/728/26 $2.00

$(1 + cv_p^2)$ and $n' = N/(1 + cv_p^2)$, where cv_p is the coefficient of variation of p (Brown et al., 1976). It is shown that apart from the simplifying assumption that stimulated release occurs over an infinitely brief time, the only necessary assumption for this is site independence, which is also a precondition for the simple binomial. Notably, if N is enlarged by any number, corresponding to hypothetical sites with $p \cong 0$, the above relationships remain true, with $\langle p \rangle$ and cv_p^2 having new values; neither mean nor variance (or overall distribution) of the outputs contains information about the number of release sites, unless the distribution of p among these sites is known. One such distribution, namely that p is the same at all "active" sites and 0 elsewhere, gives n' equal to the number of active sites, but is logically untenable (see Discussion).

One basic assumption of the binomial model, that release by a stimulus is limited to a fixed maximum, is reconcilable with reality (a noninfinitesimal time period of release) only if the release of a quantum by a site somehow entails the temporary inability to release another. As the converse of release, this "depletion" is inherently stochastic, and if subsequent "refilling" is also stochastic, events at each site constitute a Markov chain, corresponding to one of the models considered by Vere-Jones (1966) and more recently by Melkonian (1993). Here I have generalized this model in two ways, to the situation in which release sites have different output probabilities (p_o's) and refill rates, and to where these parameters vary in time, to make it possible to compute statistical outcomes for any hypothetical parameter sets and stimulus sequences, and for spontaneous release. One result that emerges is that if p_o varies between sites, cv_p and the relative contribution of quanta from different sites can be expected to change continually early in trains of stimuli, to vary with stimulation frequency, and to vary with conditions that modify p_o's or the rate of refilling; n' tends to rise with stimulation frequency and with any factor that increases p_o's.

I also present derivations of equations for the third moment of the quantal outputs, for the modification of moments by quantal amplitude variation of various types (Walmsley, 1993), and for the expected covariance between numbers of quanta released by successive stimuli, which, unlike means and variance, may contain information about whether any experimentally observed changes in $\langle p \rangle$ or cv_p might be due to changes in p_o's or p_A's (Vere-Jones, 1966). Statistical measures are also derived for "spontaneous" release.

In addition, a new function, the area product, is introduced; this is easily computed from data and provides not only the total variance of sequential synaptic signals, unbiased by recording noise, but also the distribution in time of this variance before and after the point of stimulation. With some caveats, the latter can be an indicator of whether the signals represent a mix from different synapses producing quantal responses of different time course, and/or provide an estimate of the amplitude of the channels underlying quantal responses.

METHODS

All of the equations presented were derived from basic principles using approaches found in an introductory textbook of mathematical statistics (Weatherburn, 1961) and in Vere-Jones (1966). Various calculations were done on an IBM-compatible PC, either with a spreadsheet or with programs written in C. These calculations were of two kinds: 1) verification with a Monte Carlo simulation of the formulae for variances and covariances and 2) determination of the effect of arbitrarily assigned release probabilities and replenishment rates for arrays of up to 100 release sites in which one or both of these parameters varied between sites. Some of the results of these calculations are presented in the illustrative figures in the next section. In the simulations I used, the ran2() subroutine of Press et al. (1992), which was checked to verify the absence of correlations between successive pseudorandom numbers, to obtain random numbers between 0 and 1, from which, when required, exponentially distributed random numbers could be obtained by taking the negative of the logarithm. For normally distributed random numbers I used the gasdev() subroutine of Press et al. (1992).

THEORY AND RESULTS

The simple and compound binomial distribution

Consider a synapse that has been stimulated repetitively at a constant rate, the outputs of which have settled down to a value that is constant except for statistical fluctuations. To obtain the statistical composition of the outputs (i.e., mean, variance, etc.) imagine records from a very large array of detectors that cover the whole presynaptic area. Each detector has as its territory an area so small that no more than one quantum can be released in the (supposedly) infinitely brief time period, after each stimulus, that release occurs; it signals a 1 for a "success" and a 0 for a "failure." Because the number of detectors is much larger than the number of release sites, the record from most of the detectors contains only 0's, but others contain 1's and 0's. For any one of these active sites there will be a certain number of 1's, say s, for k stimuli. The sum over the k stimuli is s, the sum of squares is s, and the sum of cubes is s. The mean $= \mu_1' =$ mean square $= \mu_2' =$ mean cube $= \mu_3' = s/k$, for which the expected value is p, the probability of release. Thus the expected value of the mean $= p$ and the second and third moments about the mean are variance $= p(1 - p)$, third moment $= p - 3p^2 + 2p^3 = p(1 - p)(1 - 2p)$, i.e., outputs from each active site are binomially distributed with parameters 1 and p.

In nearly all experimental situations the data we have for each stimulus will correspond to the sum of the outputs from the hypothetical detectors. The mean and variance (and third moment) of these are obtained simply by summation if and only if the numbers from each are uncorrelated, i.e., whether or not a success occurs at any one detector is uncorrelated with whether or not a success occurs at any others. In other words, release sites must be independent. Then,

$$E(m) = \sum p; \qquad \mathrm{var}(m) = \sum p - \sum p^2$$
$$p' \equiv 1 - \mathrm{var}(m)/E(m) = \sum p^2/\sum p$$
$$n' \equiv E(m)/p' = (\sum p)^2/\sum p^2.$$

In terms of N, the number of release sites,

$$E(m) = N\sum p/N = N\langle p\rangle \tag{1a}$$
$$\begin{aligned}\mathrm{var}(m) &= N\langle p\rangle - N(\langle p\rangle^2 + \mathrm{var}(p))\\ &= N\langle p\rangle(1 - \langle p\rangle(1 + \mathrm{var}(p)/\langle p\rangle^2))\\ &= N\langle p\rangle(1 - \langle p\rangle(1 + cv_\mathrm{p}^2))\end{aligned} \tag{1b}$$

and

$$p' = \langle p\rangle(1 + cv_\mathrm{p}^2) \tag{1c}$$
$$n' = N/(1 + cv_\mathrm{p}^2). \tag{1d}$$

This is the "compound binomial." Note that Eqns. 1 are valid for any assumed value of $N \geq n'$ (because $cv_\mathrm{p}^2 \geq 0$); the mean and variance of outputs can give true N only if cv_p is known a priori, or give true cv_p only if true N is known a priori. For one particular distribution of p, where p is the same at all active sites and zero elsewhere, the situation is the same as if the silent sites did not exist and in a sense the simple binomial ($cv_\mathrm{p} = 0$) pertains; n' is the number of active sites, invariant with any alteration of p as long as $cv_\mathrm{p} = 0$, but true N remains unknown. Indeed, we could not determine true N even if we had access to the records from each and every detector, because a record with no successes does not preclude a past or future success.

It must be emphasized that if sites are not independent, Eqns. 1 do not pertain, even for $cv_\mathrm{p} = 0$, which is otherwise the simple binomial (Brown et al., 1976).

Comparison of output distributions for simple and compound binomials

The third moment of the outputs is given by

$$\begin{aligned}M3 &= \sum p - 3\sum p^2 + 2\sum p^3\\ &= N\langle p\rangle - 3N[\langle p\rangle^2 + \mathrm{var}(p)]\\ &\quad + 2N[\langle p\rangle^3 + 3\langle p\rangle\,\mathrm{var}(p) + P3]\\ &= N[\langle p\rangle(1 - \langle p\rangle)(1 - 2\langle p\rangle)\\ &\quad - 3\,\mathrm{var}(p)(1 - 2\langle p\rangle) + 2P3]\end{aligned} \tag{1e}$$

where *M3* and *P3* denote the third moments of m and p about their means. In general *M3* is not the same for a compound binomial as for the simple binomial, and it might therefore be supposed that from the output distribution one could determine whether a simple or a compound binomial pertains. However, simulations of outputs for a compound binomial show that this is not the case (Brown et al., 1976). The examples in Table 1 are for arrays of 100 sites with arbitrary widely varying p. In each case the first column is a list of the number of outputs with 0, 1, 2, etc. quanta expected for 1000 iterations. The second and third columns are calculated using the mean ($\langle m\rangle$) and variance of the outputs to obtain p' and n', rounding off n' down and up to integers and choosing for each a new $p' = \langle m\rangle/(\mathrm{new})\ n'$. For example, in the first set, certain parameters producing $\langle m\rangle = 1$ gave $p' = 0.30$ and $n' = 3.34$; the two simple binomials for comparison have $n = 3$ and $p = 0.33$ and $n = 4$ and $p = 0.25$. The true distribution is either between the two for simple binomials, or different from an extreme of the latter by no more than 4, i.e., not statistically significant; the same was true for any set of p's giving about 30% or more "failures." Apart from the appearance of a few outputs more than n', appreciable differences between the compound binomial and the corresponding simple binomials occur only if p' is more than about 0.5 and $\langle m\rangle$ is so high that there are no failures and few if any unitary responses. The corollary is that whereas an output distribution may sometimes show that a simple binomial does not pertain, the absence of significant deviations from the simple binomial is insufficient to deny a compound binomial. It may indeed be shown explicitly that the probabilities of a failure and of a single unit response are for any mixture of p's indistinguishable from the corresponding simple binomial, provided no single p is more than about 0.3 (see Appendix, 1). Virtually the only information on the output distribution of p's at individual sites is that none of them can be more than the fraction of "successes."

Composition of p

Let us suppose that a release site when stimulated may or may not be able to release a quantum; one precondition for capability might be the presence of an available quantum, i.e., its being "filled", and for simplicity of expression let us suppose that this is the case. Then, its p will be the product of p_o (the chance it releases if it has an "available" quantum) and p_A (the chance that a quantum is available), i.e., $p = p_\mathrm{o}p_\mathrm{A}$. The logic leading to Eqns. 1 remains unchanged, and it follows that there is no way of determining from an output distribution (mean, variance, etc.) either $N\langle p_\mathrm{A}\rangle$, the mean number of sites capable of release, or $\langle p_\mathrm{o}\rangle$, the mean probability of release of filled sites, even if N is known a priori, or if the simple binomial pertains. Correspondingly, any experimentally induced change in $\langle m\rangle$ (quantal content) might be due to a change in $\langle p_\mathrm{A}\rangle$ and/or $\langle p_\mathrm{o}\rangle$. Furthermore, if we had detectors of "quantal availability" at each site, these would each produce a succession of 0's and 1's, by definition binomially distributed.

TABLE 1 Distributions of quanta released per stimulus for compound binomials

	$\langle m \rangle = 1.0$			$\langle m \rangle = 2.0$			$\langle m \rangle = 4.0$		
	$n' = 3.34$	3	4	$n' = 3.83$	3	4	$n' = 5.33$	5	6
	$p' = 0.30$	0.33	0.25	$p' = 0.52$	0.67	0.5	$p' = 0.75$	0.8	0.667
0	299	296	316	46	37	62	0	0	0
1	448	444	422	*262	221	249	*2	6	17
2	209	222	211	402	444	375	51	51	82
3	40	37	47	*228	298	251	*258	205	220
4	4	0	4	55	0	63	396	410	329
5	0	0	0	*7	0	0	*230	328	263
6	0	0	0	0	0	0	57	0	88
7	0	0	0	0	0	0	*6	0	0
	$\langle m \rangle = 4.0$			$\langle m \rangle = 7.05$			$\langle m \rangle = 8.0$		
	$n' = 12.8$	12	13	$n' = 12.0$	11	12	$n' = 10.8$	10	11
	$p' = 0.31$	0.33	0.31	$p' = 0.59$	0.64	0.59	$p' = 0.74$	0.8	0.73
0	7	8	8	0	0	0	0	0	0
1	46	46	48	0	0	0	0	0	0
2	129	126	128	2	2	3	0	0	0
3	213	211	210	12	12	15	0	1	2
4	237	238	234	48	43	49	4	6	10
5	187	192	188	*118	107	113	30	27	39
6	110	112	112	*198	191	187	*109	88	103
7	49	48	50	233	244	229	*226	202	197
8	17	15	17	*195	218	204	277	302	262
9	5	3	4	*119	130	129	*211	268	233
10	1	0	1	53	46	55	*103	107	124
11	0	0	0	*17	8	14	*33	0	30
12	0	0	0	*4	0	2	*7	0	0
13	0	0	0	*1	0	0	*1	0	0

Expected numbers for 1000 stimuli, from arrays of 100 sites. In each case the first column gives the number expected for the compound binomial and the second and third columns give the numbers expected for apparent n (n') rounded down and up to the closest integers.

*Values where with sufficient iteration, there might be a significant difference between numbers of quanta observed with the compound binomial and numbers predicted by simple binomials.

Temporal fluctuations in p

In reality p_o at each site and therefore p must be a temporally random variable, because it depends upon the stochastic opening of voltage-gated Ca^{2+} channels, the stochastic closing of these channels, the stochastic combination of intracellular Ca^{2+} with its receptor(s), etc., which cannot be identical from stimulus to stimulus. From the derivation of Eq. 1, this has no effect whatever on the statistical outcome.

Summary of assumptions for binomial distribution

In the above derivation there are a number of explicit and implicit assumptions:

1. A release site is capable of releasing only one quantum per stimulus. This has been implicit in all statistical treatments of transmitter release; the Poisson distribution at frog neuromuscular junction in low Ca^{2+}/raised Mg^{2+} (del Castillo and Katz, 1954a) is incompatible with single release sites releasing more than one quantum at a time. In principle, if any release site *could* release more than one quantum, it would have to be considered as two or more sites that are possibly linked and correlated.

2. The independence of sites, in the sense that fluctuations of release from one site to another are uncorrelated, has also been implicit in all statistical treatments of release. If correlations do exist, positive correlations increase variance, whereas negative correlations reduce variance (see Dityatev et al., 1992), for both the simple and compound binomial.

3. The assumption of stationarity is unnecessary, provided any nonstationarity of $p = p_o p_A$ at any site is uncorrelated with nonstationarity at any other. If p's vary in tandem over time, corresponding to the usual definition of nonstationarity, this gives a positive correlation and an increase in variance.

4. The hidden assumption of depletion is necessary to reconcile assumption 1 and a finite time course of release. It and its ramifications will be considered in detail below.

Depletion and refilling: statistical measures in trains

In the classic review by del Castillo and Katz (1956) it was pointed out that the end-plate potential (EPP) produced by a nerve impulse represents an intense transient acceleration of "spontaneous" quantal release, and there is no reason to doubt that this is generally the case. The time of this transient, although brief, cannot be instantaneous, and therefore the probability of release from any site can in principle be expressed as a series of very small probabilities in very small time periods (δt). It follows that if site capability were to remain unaffected after quantal discharge, release in each δt would be Poisson distributed, and net release would be unlimited and Poisson. Thus a binomial model depends on the assumption that the release of a quantum from a site by a stimulus entails no further release by the same stimulus from the same site. The incapability of the site to release a second quantum can be termed "depletion," and recovery, for a subsequent stimulus, can be termed "replenishment" or "refilling," without necessarily implying that these processes physically represent the loss of a preformed quantum and acquisition by the site of a new preformed quantum.

In the absence of any reason to believe the contrary, namely, replenishment at a fixed time after a quantum is released, refilling is hereafter assumed to be a stochastic process that is not necessarily complete between stimuli. As will be seen below, this depletion model provides a rationale for the analysis of data, particularly from short trains of stimuli, to obtain some insight into whether experimentally produced changes in $\langle m \rangle$ reflect change in p_o's or p_A's.

Dependence of p_A on p_o and statistics of outputs in trains

Vere-Jones (1966) has rigorously derived the statistical makeup of outputs produced by a series of stimuli, for the case in which there is a constant probability of release (p_o) from n equivalent sites, each either filled or unfilled, with constant probability (α) of unfilled sites becoming refilled between stimuli. This model constitutes a "simple, discrete-time, positive recurrent Markov chain." To summarize his result, $\langle m \rangle$ and var(m) tend geometrically to equilibrium values np and $np(1 - p)$, respectively, with $p = p_o p_A$ and $p_A = \alpha/(1 - q_o\beta)$, β being $(1 - \alpha)$ and q_o being $(1 - p_o)$. The number of quanta available for release (np_A) is always binomially distributed and positively correlated from stimulus to stimulus; outputs are binomially distributed, and the correlation between successive outputs is always negative. In particular, at equilibrium the covariance of successive outputs is $p_o^2 q_o \beta(\mathrm{var}(n) - E(n)) = -np_o^2 q_o \beta$. The geometric progression to equilibrium, which depends upon constant p_o and α, was known even then to be an oversimplification, in view of the data of Elmqvist and Quastel (1965).

The logic of Vere-Jones (1966) gives rise to a fairly simple algorithm that makes it possible to obtain solutions, in terms of probabilistic outcomes, for any set of release probabilities and replenishment probabilities at any array of release sites, and for any stimulation paradigm, without recourse to Monte Carlo simulation, as follows.

Consider a group of n sites with identical p_o and α, and designate as $\mathrm{n_i}$ the number of sites with a quantum available for release, i.e., the number of filled or eligible release sites, at the ith stimulus, for which p_o is p_i, α is α_i (and $\beta_i = 1 - \alpha_i$). In general, $\mathrm{n_i}$ is a random variable with a mean (say n) and a variance equal to $\mathrm{n}(1 - \mathrm{n}/n)$, because n is binomially distributed. One can keep track of what occurs with each stimulus and, subsequently, with the following general scheme:

At the ith stimulus $\mathrm{n_i} = \mathrm{n} + \epsilon_1$; $E(\mathrm{n_i}) = \mathrm{n}$
quanta released $= m_i = p_i(\mathrm{n} + \epsilon_1) + \epsilon_2$; $E(m_i) = p_i E(\mathrm{n_i})$
filled sites remaining $= f_i = q_i(\mathrm{n} + \epsilon_1) - \epsilon_2$
unfilled sites $= u_i = n - q_i(\mathrm{n} + \epsilon_1) + \epsilon_2$
unfilled sites after partial refill $= v_i = \beta_i(n - q_i(\mathrm{n} + \epsilon_1)$
$+ \epsilon_2) + \epsilon_3$
filled sites after partial refill $= \mathrm{n_{i+1}} = \alpha_i n + \beta_i q_i(\mathrm{n} + \epsilon_1)$
$- \beta_i \epsilon_2 - \epsilon_3$
quanta released $= m_{i+1} = p_{i+1}(\alpha_i n + \beta_i q_i(\mathrm{n} + \epsilon_1) - \beta_i \epsilon_2$
$- \epsilon_3) + \epsilon_4$

Here the ϵ's designate independent "error" terms; each has an expected value of 0 and an expected value of the square (or cube) that accords with the binomial sampling that generates it, e.g., $\epsilon_2{}^2 = \mathrm{n} p_i q_i$. Expected values for m_i, f_i, etc. are given by the expressions with error terms omitted. In a numerical Monte Carlo simulation, each ϵ occurs where a decision is made according to the value of a random number. The variance at each stage can be obtained by squaring the expression containing one or more ϵ's and retaining only terms that include squares of ϵ's; the variance obtained in this way will be the same as that deduced from the binomial distribution of each, with parameter n:

$$\varepsilon_1^2 = \mathrm{var}(\mathrm{n_i}) = E(\mathrm{n_i})(1 - E(\mathrm{n_i})/n)$$

$$\varepsilon_2^2 = p_i q_i E(\mathrm{n_i})$$

$$E(m_i) = p_i E(\mathrm{n_i}) = p_i E(\mathrm{n_i}) = np_i p_{Ai} \tag{2a}$$

$$\begin{aligned} \mathrm{var}(m_i) &= p_i^2 \varepsilon_1^2 + \varepsilon_2^2 \\ &= p_i^2\, \mathrm{var}(\mathrm{n_i}) + p_i q_i E(\mathrm{n_i}) \\ &= E(m_i)(1 - E(m_i)/n) \\ &= np_i p_{Ai}(1 - p_i p_{Ai}) \end{aligned} \tag{2b}$$

$$\varepsilon_3^2 = \alpha_i \beta_i (n - q_i E(\mathrm{n_i}))$$

$$\varepsilon_4^2 = p_{i+1} q_{i+1} E(\mathrm{n_{i+1}})$$

$$E(\mathrm{n_{i+1}}) = \alpha_i n + \beta_i q_i E(\mathrm{n_i})$$

$$E(m_{i+1}) = p_{i+1} E(\mathrm{n_{i+1}})$$

$$\mathrm{var}(m_{i+1}) = E(m_{i+1})(1 - E(m_{i+1})/n),$$

where p_{Ai} has been written for $E(\mathrm{n_i})/n$, the probability that a site has a quantum "available" at the ith stimulus. The covariance between any two stages is given by multiplying the two relevant expressions and omitting terms that are not squares of ϵ's:

$$\mathrm{cov}(\mathrm{n}_i, \mathrm{n}_{i+1}) = \varepsilon_1^2 \beta_i q_i$$

$$\begin{aligned}\mathrm{cov}(m_i, m_{i+1}) &= p_{i+1}\beta_i(p_i q_i \varepsilon_i^2 - \varepsilon_2^2)\\ &= p_{i+1}\beta_i p_i q_i(\mathrm{var}(\mathrm{n}_i) - E(\mathrm{n}_i))\\ &= -np_{i+1}\beta_i p_i q_i p_{Ai}^2 \qquad (2c)\end{aligned}$$

Equation 2a can also be written in terms of $E(m_i)$ and $E(m_{i+1})$,

$$\mathrm{cov}(m_i, m_{i+1}) = E(m_i)(p_{i+1}\alpha_i - E(m_{i+1})/n) \qquad (2d)$$

Following the logic to the next and subsequent stimuli, we find that, in general,

$$\mathrm{cov}(m_i, m_{i+k}) = \mathrm{cov}(m_i, m_{i+k-1})\beta_{i+k-1} q_{i+k-1} p_{i+k}/p_{i+k-1} \qquad (2e)$$

From the above formulae, given any sequence of p_o's and α's, one can list expected values of numbers of filled (occupied/capable/eligible) sites, expected values of outputs ($\langle m \rangle$), variances, and covariances, from which one can also obtain these measures for sums of outputs over time. This permits calculation not only of responses to iterated trains of stimuli (where facilitation might change p_o's and/or α's), but also of what happens if release by each stimulus is dispersed in time, i.e., each p_o is replaced by a series of small p_o's in small time bins after each stimulus (see Release Asynchrony below), and what happens with continuous "spontaneous" release. The covariances are essential for the variance of summed outputs. With application to trains, variances and covariances of course pertain to m's at times where the expected values of m are the same, e.g., variance between the number 3's of repeated trains, covariance between seconds and firsts.

For arrays of N independent sites with different p_o's and/or α's, one uses the above equations, with $n = 1$, for each site, and adds all means, variances, and covariances to obtain values for the whole array. Such summation gives the usual expressions for the compound binomial: $\langle m \rangle$ (or $E(m)$) $= N\langle p \rangle$ and $\mathrm{var}(m) = N\langle p \rangle (1 - \langle p \rangle (1 + cv_p^2))$, where each p, at each site and at each stimulus, is its $p_i p_{Ai}$ ($p_o p_A$ at the ith stimulus), but summed covariances cannot be expressed in a neat mathematical expression. Notably, because each q_o appears separately from $p_{i+1} p_i p_{Ai}^2$ in Eq. 2c, covariances, as well as the progression of $\langle m \rangle$'s, contain information on p_o's.

Parameters α and p_o are of course always ≤ 1. In setting up models it is convenient to define a parameter R_A (refill rate, ≥ 0) and to set $\alpha = 1 - \exp(-R_A T)$, where T is the time between stimuli, so that R_A's may be modified freely without α's becoming more than 1. Similarly, p_o must always be less than 1, and it is convenient (also see below) to define a parameter r, with $p_o = 1 - \exp(-r)$, that can be postulated to rise to any arbitrary extent during a stimulus train (or with increase in [Ca^{2+}]). Reasonable values for these parameters can be assessed only roughly from available data. From Elmqvist and Quastel (1965), for the human neuromuscular junction $\langle p_o \rangle$ (estimated from "rundown," weighted by contribution to EPP, in curare and normal $\mathrm{Ca}^{2+}/\mathrm{Mg}^{2+}$) varies considerably between junctions but averages roughly 0.3 at the start of trains and then grows, depending on stimulation frequency. Mean R_A starts at about 1.5/s, but can grow with high-frequency stimulation to about 10/s; at 100 Hz $\langle p_A \rangle$ is probably about 0.1 at about the 40th stimulus in a train and thereafter slowly declines. If these estimates are even roughly correct, the nearly constant variance/mean for EPPs, over a wide range of stimulus frequencies, is incompatible with anything but stochastic replenishment. Mennerick and Zorumski (1995) give 380 ms for the time constant for recovery from paired pulse depression for cultured hippocampal EPSCs, i.e., $R_A \approx 2.6$/s, and from paired pulse depression, $\langle p_o \rangle$ (again weighted by contribution to EPSC) varies widely but is often about 0.5. From data on arthropod neuromuscular junctions (see McLachlan, 1978), initial $\langle p_o \rangle$'s are much lower and facilitation (? rise in p_o's) is very prominent.

Examples of hypothetical outputs during trains

With $n = 1$, $E(\mathrm{n}_i)$ is p_{Ai} and $E(m_i)$ is $p_i E(\mathrm{n}_i)$. Starting with, for example, $p_{A1} = 1$ at each site, to obtain the succession of p_{Ai}'s at each site as the train progresses, the required equation is merely

$$E(\mathrm{n}_{i+1}) = p_{A,i+1} = \alpha_i + \beta_i q_i E(\mathrm{n}_i).$$

If refill between trains is incomplete, one uses at the end of each train the α for the intertrain interval to obtain a new p_{A1} and repeats the whole sequence until p_{A1}'s no longer change. Then,

$$E(m_i) = p = p_i E(\mathrm{n}_i); \qquad \mathrm{var}(m_i) = E(m_i)(1 - E(m_i));$$

$$\mathrm{cov}(m_i, m_{i+1}) = E(m_i)(p_{i+1}\alpha_i - E(m_{i+1})).$$

For the whole array one sums over all N sites to obtain $E(m) = \Sigma\, p = N\langle p \rangle$ and $\mathrm{var}(m) = \Sigma\, p - \Sigma\, p^2 = N\langle p \rangle (1 - \langle p \rangle (1 + cv_p^2))$ and $\mathrm{cov}(m_i, m_{i+1})$ for each stimulus.

Using a spreadsheet and these equations for an array of N from 2 to 100 sites with varied initial p_o, and given a tendency for r (and therefore p_o) to rise asymptotically, I find that $\langle m \rangle$ never falls exponentially but may rise or fall monotonically, or rise and subsequently fall, or fall and then rise and fall again, depending upon the parameters introduced; this occurs because outputs from initially high p_o sites run down, whereas outputs from low p_o sites run up to an equilibrium, all at different rates.

Fig. 1 *A* shows how p_A's and p's (each $= p_o p_A$) evolve during a train of stimuli, for a model, drastically simplified for illustrative purposes, with only two kinds of sites and constant p_o's: $p_o = 0.8$ at 20 "high-p" sites and 0.08 at 80

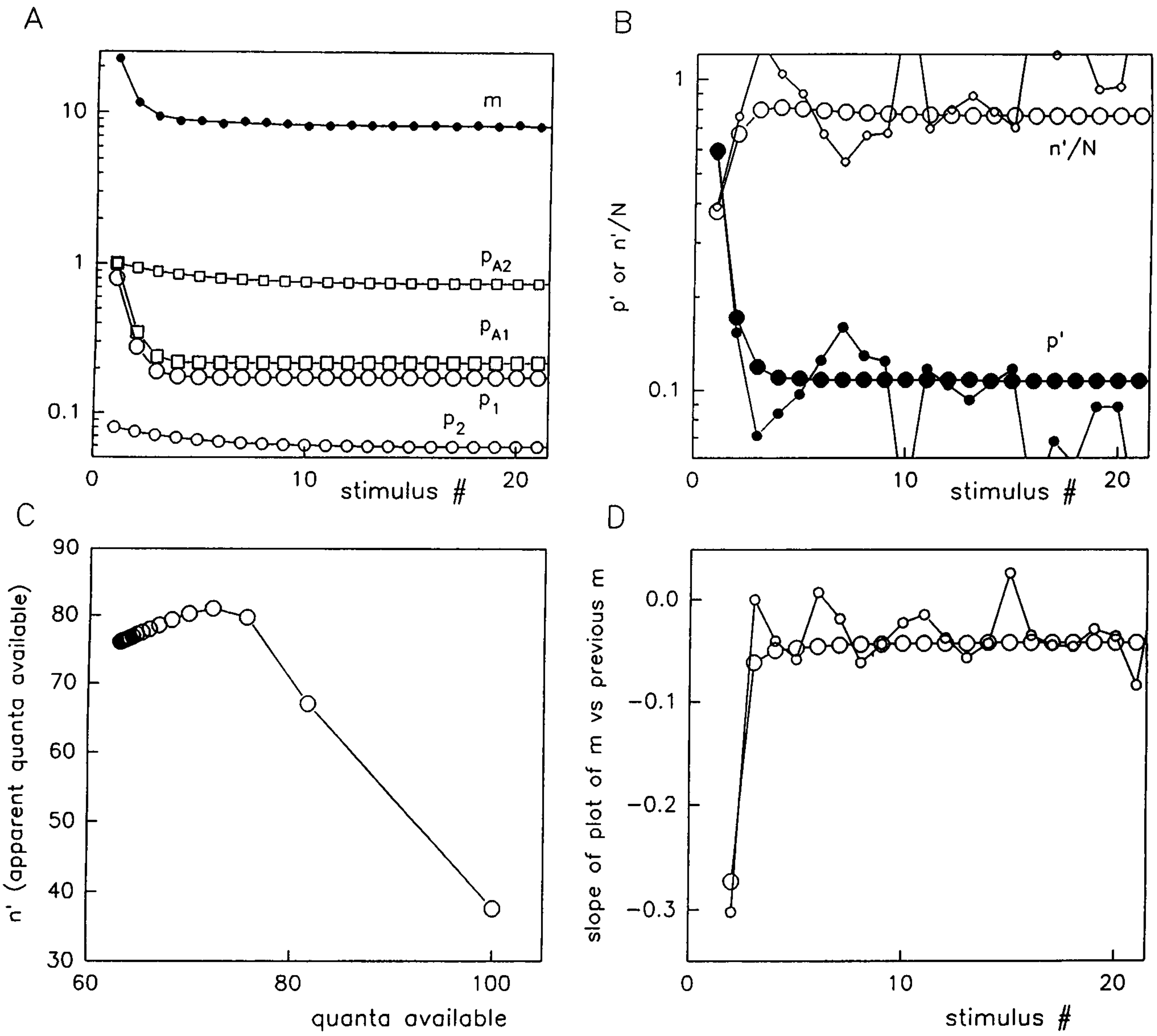

FIGURE 1 Evolution of outputs, p_A, p, and n' during a train of stimuli for an array with 20 "high p" and 80 "low p" sites, with p_o = 0.8 and 0.08, respectively. The stimulus frequency is 25 Hz and R_A is 5/s, giving α = 0.181 at all sites. The time interval between trains is 2 s, giving nearly complete "refill." As the train proceeds, p_A, and therefore p (= $p_o p_A$), falls faster and to a greater extent at "high p" sites—p_1 and p_2 in *A* are $p_o p_A$ at the two kinds of sites—resulting in a fall in p' and an increase in n' (*B*). (*C*) Counterintuitive relation between n' and the actual number of quanta available, $N\langle p_A \rangle$. The correlation between m's at successive stimuli is shown in *D*. In *B* and *D* the small symbols are values realized by a Monte Carlo simulation with 1000 trains. They illustrate the rather high sampling error of p' and n'. Lines merely join points.

"low-p" sites with the same constant R_A at both sites, giving α = 0.181 for the interstimulus interval of 40 ms (25 Hz), and nearly complete refill in the 2-s between-train interval. As the train progresses to equilibrium, p_A declines more and faster at the high-p sites (Fig. 1 *A*), resulting in a progressively reduced cv_p, and therefore reduced p' and increased n' (Fig. 1 *B*); the small symbols in Fig. 1 *B* are from a Monte Carlo simulation with 1000 trains. The theoretical relation between n' and the actual number of available quanta ($N\langle p_A \rangle$) is counterintuitive—n' rises as the number of filled sites decreases (Fig. 1 *C*). The slope of a plot of output versus previous output (cov(m_i, m_{i+1})/var(m_i) in Fig. 1 *D*) is substantially negative only for the first pair in the train—simulation-realized values (*small symbols*) are close to the theoretical.

Fig. 2 shows similar calculations for the same model, but with hypothetical facilitation, such that r's—each p_o is 1 − exp(−r)—increase throughout the train (eventually by ninefold). At the "low-p" sites, p_o increases from 0.08 to 0.53, but at the "high-p" sites p_o rises merely from 0.8 to 0.9999. This rather high overall facilitation is hardly manifest, except as a reduction of how much $\langle m \rangle$ falls and a small increase in the extent to which n' rises, because increases in p_o's are counterbalanced by falls in p_A's.

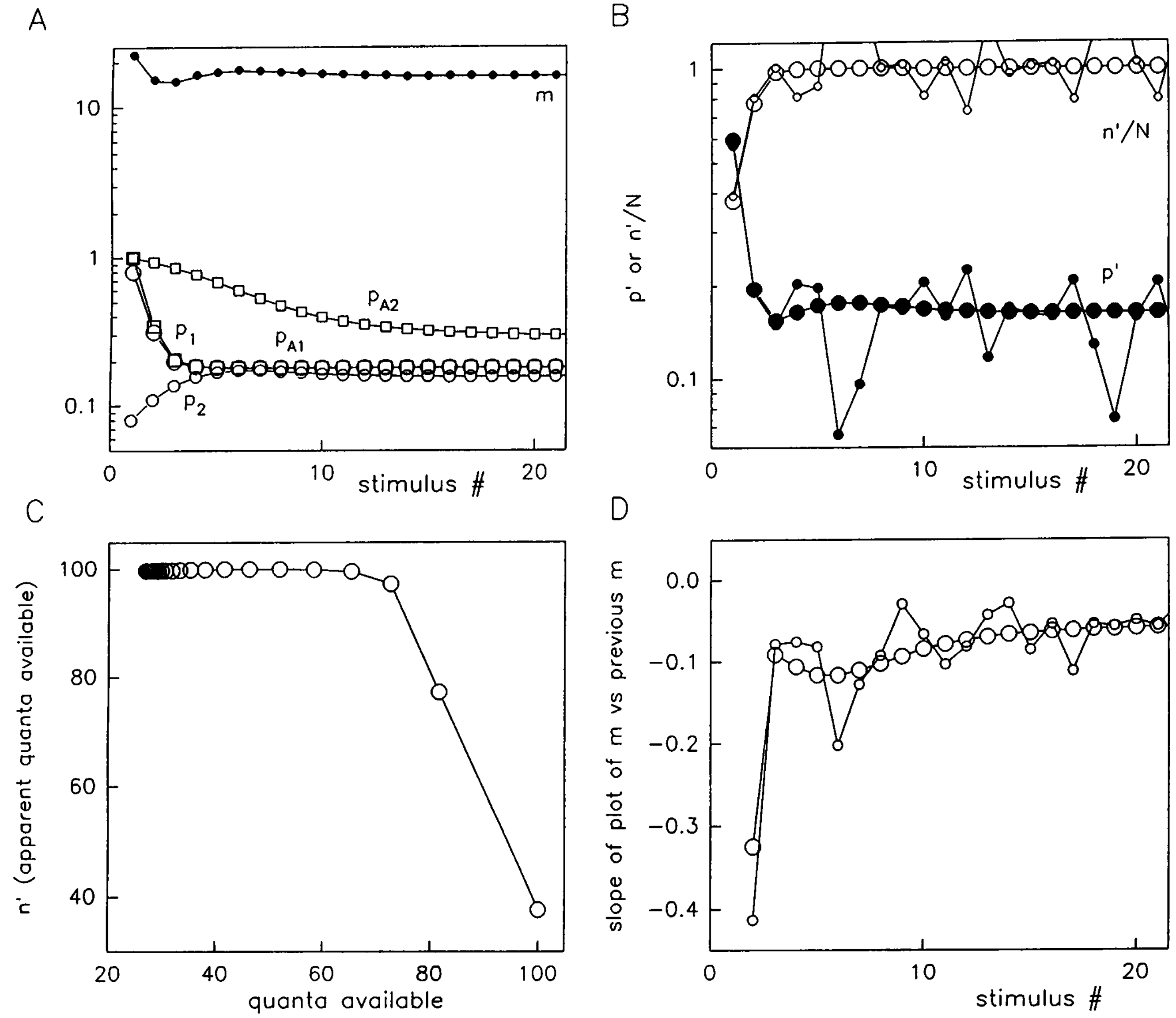

FIGURE 2 Same as Fig. 1, except for the addition of facilitation such that r ($p_o = 1 - e^{-r}$) grows with each stimulus, to a maximum of ninefold, at both kinds of site. Overall mean m rises after an initial fall and is maintained higher than without facilitation (Fig. 1), and n'/N approaches unity (B) as cv_p becomes small. Note that the evolution of p' and mean m gives little hint of the rise in p_o.

In Fig. 3 are shown plots from Monte Carlo simulations, of the second m (m_2) versus the first (m_1) with 200 or 5000 stimulus pairs, here with one "high p" (p_o = 0.8) and 4 "low p" (p_o = 0.08) sites, either with p_o's unchanged at the second stimulus (Fig. 3 A) or with facilitation made so large (3.2-fold multiplication of r's at the second stimulus) that m_2 is larger than m_1 (Fig. 3 B). Note: 1) the substantial number of occasions where output is higher than n' (for m_1, theoretical 1.885, 1.66 in A, 1.83 in B, for 200 trains); 2) depletion is not necessarily signaled by a decrease in m ("paired pulse depression"), because it can be counterbalanced by facilitation of p_o, but is always signaled by a negative correlation between m_2 and m_1; and 3) 200 stimulus pairs have been sufficient to show significant negative correlation between m_2 and m_1. Similar simulations with high N (not illustrated) gave essentially linear relations between m_2 and m_1 and significant correlation (for 200 iterations) whenever the absolute value of the theoretical slope (cov(m_2, m_1)/var(m_1)) was more than about 0.2, and always if all sites start (nearly) full and p_o's at some sites are high enough that outputs fall at the second stimulus. This agrees with the highly significant negative correlation between second and first EPPs in trains reported by Elmqvist and Quastel (1965) in normal Ca^{2+}/Mg^{2+} and curare, and the lack of correlation between EPP pairs in low Ca^{2+}/high Mg^{2+} (del Castillo and Katz, 1954b).

A variation of the theme in Figs. 1 and 2 is given in Fig. 4, where the postulate is that stimuli have been given in the

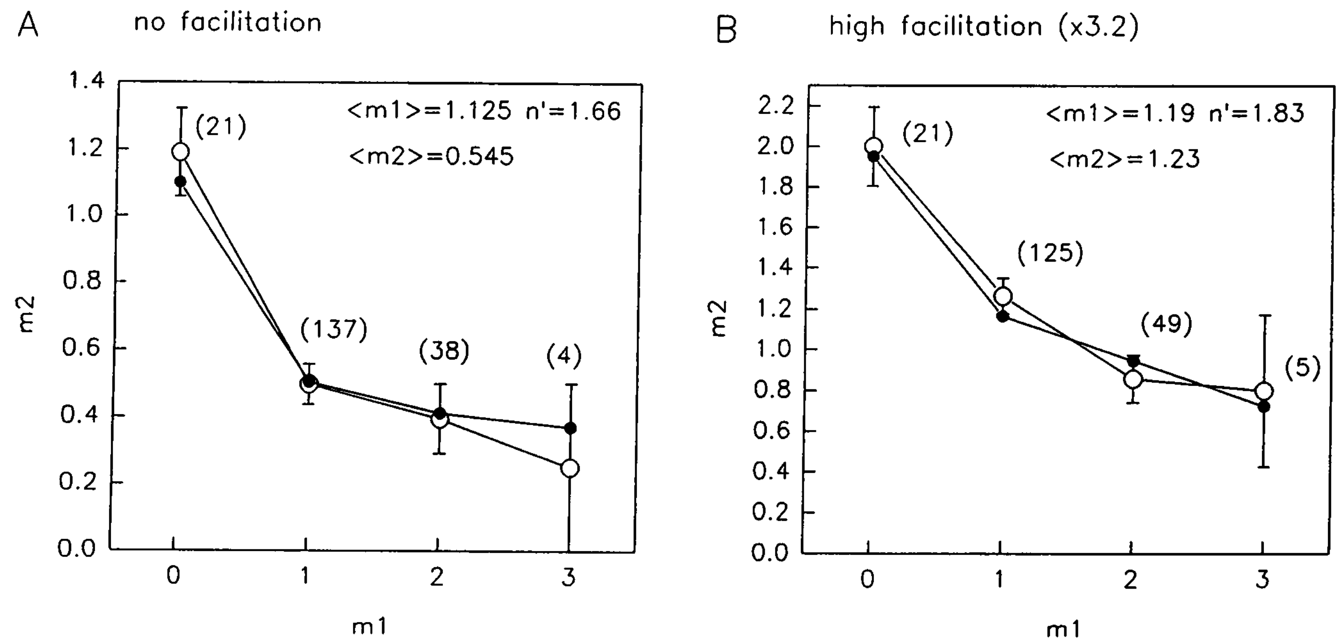

FIGURE 3 Plots of m at a second stimulus (m2) versus m at a first stimulus (m1) for paired pulses; here the simplified array has one "high p" and four "low p" sites, with initial p_o of 0.8 and 0.08. Monte Carlo simulations were done with 200 pairs (*large open circles*, number of times value of m1 realized in brackets) or 5000 pairs (*small filled circles*, bars within points). In *A*, p_o are unchanged at the second stimulus, whereas in *B*, r are facilitated 3.2-fold, giving m2 somewhat bigger than m1. Note that 200 pairs are sufficient to show nonindependence of outputs, resulting from depletion and incomplete refill between the two stimuli.

presence of Sr^{2+} (Bain and Quastel, 1992a), producing after each stimulus a "tail" of release generated by residual Sr^{2+} in the nerve terminal. The relative contributions of outputs from low- and high-p sites vary in time, because high-p_o sites become more depleted and are less likely to have a quantum available to be released by the residual Sr^{2+} if, as here, R_A is the same for both types of site.

Not illustrated here is the correlation of outputs, at successive stimuli, with the sum of previous outputs. In the simplest case (complete refill before the train, no refill between stimuli and uniform p_o's) successive outputs are $E(m_1) = Np_{o1}$, $E(m_2) = Np_{o2}(1 - p_{o1})$, $E(m_3) = Np_{o3}(1 - p_{o2})(1 - p_{o1})$, etc. The variance of sums is given simply by $N \Sigma E(m)(1 - \Sigma E(m))$, and the expected slope of m_k versus $(m_1 + m_2 + \cdots + m_{k-1})$ is simply p_{ok}, p_o at the kth stimulus. However, with an array with varied p_o's and some refill between stimuli, the most that can be said is that the above slope roughly approximates $\langle p_{ok}\rangle(1 + cv^2_{pok})\beta^{k-1}$ for the first few stimuli in the train—the correlations are as readily detected as that between m_1 and m_2—provided refill is nearly complete between trains.

Statistics of equilibrium outputs

With a continued train of stimuli, we can expect to reach an equilibrium in the sense that p_o's, α's, and expected values of n_i no longer change, i.e., all α_i's are α, all p_i's are p_o, and $E(n_{i+1}) = E(n_i) = E(n)$—the assumption is that (average) release is balanced by (average) replenishment. At this equilibrium, for n equivalent sites,

$$E(\mathrm{n}) = \alpha n + \beta q_o E(\mathrm{n}) = \alpha n/(1 - \beta q_o)$$

$$p_A \equiv E(\mathrm{n})/n = \alpha/(1 - \beta q_o) = \alpha/(\alpha + p_o - \alpha p_o) \qquad (3a)$$

$$E(\mathrm{n}) = np_A$$

$$\mathrm{var}(\mathrm{n}) = np_A(1 - p_A).$$

Writing p for $p_o p_A$,

$$E(m) = np \qquad (3b)$$

$$\mathrm{var}(m) = np(1 - p) \qquad (3c)$$

$$\mathrm{cov}(m_i, m_{i+1}) = -np^2\beta q_o[=\beta p_o^2 q_o(\mathrm{var(n)} - \mathrm{E(n)})] \qquad (3d)$$

$$\mathrm{cov}(m_i, m_{i+k}) = -np^2(\beta q_o)^k. \qquad (3e)$$

The formulae for p_A, variance, and $\mathrm{cov}(m_i, m_{i+1})$ are the same as rigorously derived by Vere-Jones (1966). As previously, one sets $n = 1$ for each site and sums over all N sites to obtain the mean, variance, and covariance for the whole array, obtaining the usual expressions for the compound binomial: $\langle m\rangle = N\langle p\rangle$ and $\mathrm{var}(m) = N\langle p\rangle\,(1 - \langle p\rangle\,(1 + cv_p^2))$, where each p is $p_o p_A$. Although the covariances contain information on p_o's, it may be noted that for a single site the negative of the ratio $\mathrm{cov}(m_i, m_{i+1})/\mathrm{var}(m)$ is $\beta p_o q_o/(q_o + p_o\beta/\alpha)$, which has a maximum of 0.125 at $p_o = q_o = \alpha = \beta = 0.5$, and at these values the covariance decreases

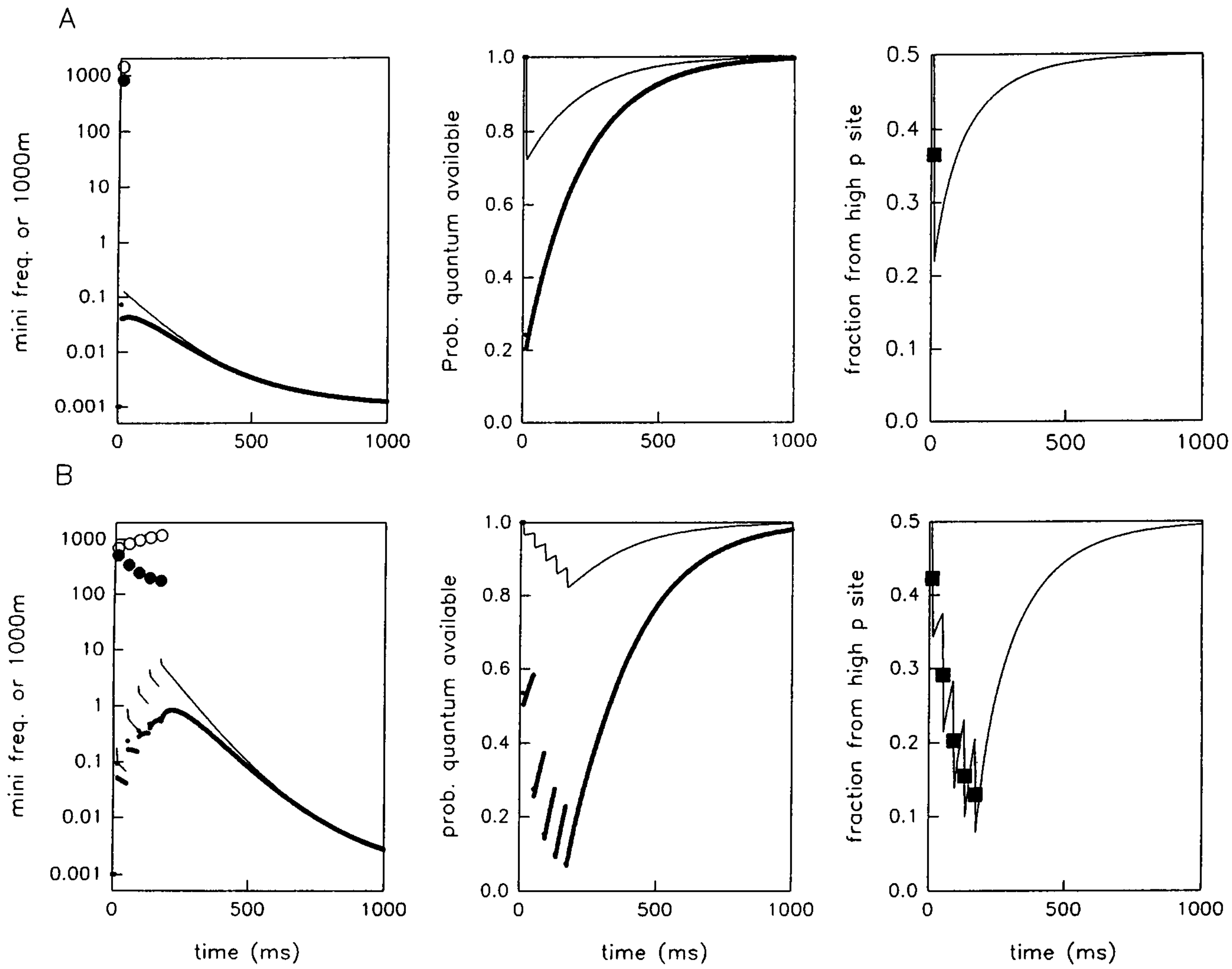

FIGURE 4 Theoretical outputs with stimulation in the presence of Sr^{2+}, with one high-p site and five low-p sites and initial R_0 such that a half "resting" miniature frequency is from the high p site. At the high p site, higher output from the stimulus (80% of the stimulus available in *A*, 50% in *B*) produces more depletion (second graph), causing the tail of raised miniature frequency for the high p site to be less than the combined tail from the low p sites. In the first two graphs filled points and heavy lines pertain to the high p site. After one stimulus (*A*) or a series of stimuli (*B*), the fraction of total output from the high p site (third graph, *filled squares* are fractions of $\langle m \rangle$) is lowered. In *B* (four stimuli) it is assumed that $[Sr^{2+}]$ is less than in *A*, so that per-pulse m and depletions are less. The time constant for the removal of intracellular Sr^{2+} is assumed to be 200 ms.

by 75% for each subsequent stimulus; at equilibrium, $\mathrm{cov}(m_i, m_{i+k})$ is unlikely ever to be detectable for $k \geq 2$.

"Automatic" changes in p' and n' with stimulation frequency

Theoretical equilibrium situations for a range of stimulus frequencies are illustrated in Fig. 5. Here an array of 100 sites has r's varying over a 1000-fold range. R_A's were randomly assigned, with an exponential distribution with a mean of 5/s. Three scenarios are shown: 1) r's (and p_o's) and R_A's remain constant; 2) r's increase exponentially with stimulus frequency, 10-fold at 50 Hz and 100-fold at 100 Hz, but R_A's remain constant; and 3) r's increase as in 2), but R_A's grow in proportion to stimulus frequency, so that α's are constant. Notably, unless R_A's rise with stimulus frequency (3), potentiation of r's (2) is scarcely manifest in net outputs (Fig. 5 *A*), because high p_o's become associated with depletion (Fig. 5 *B*). The apparent number of quanta available (n') always increases with stimulation frequency (Fig. 5 *B*), whether or not the true number of quanta available ($N\langle p_A \rangle$) falls appreciably. Why this should be is shown in the cumulative distributions for the 100 sites of p ($p_o p_A$) in Fig. 5, *D–F*. At 1 Hz the distributions are nearly the same for all three scenarios (Fig. 5 *D*), and p's are very widely distributed—a few can be near 1 because refill is nearly complete between stimuli, but at 50 or 100 Hz (Fig. 5, *E* and *F*), with or without potentiation of r's, the distributions always narrow (cv_p diminishes and n' increases), either because refill is incomplete and high-p_o sites develop low

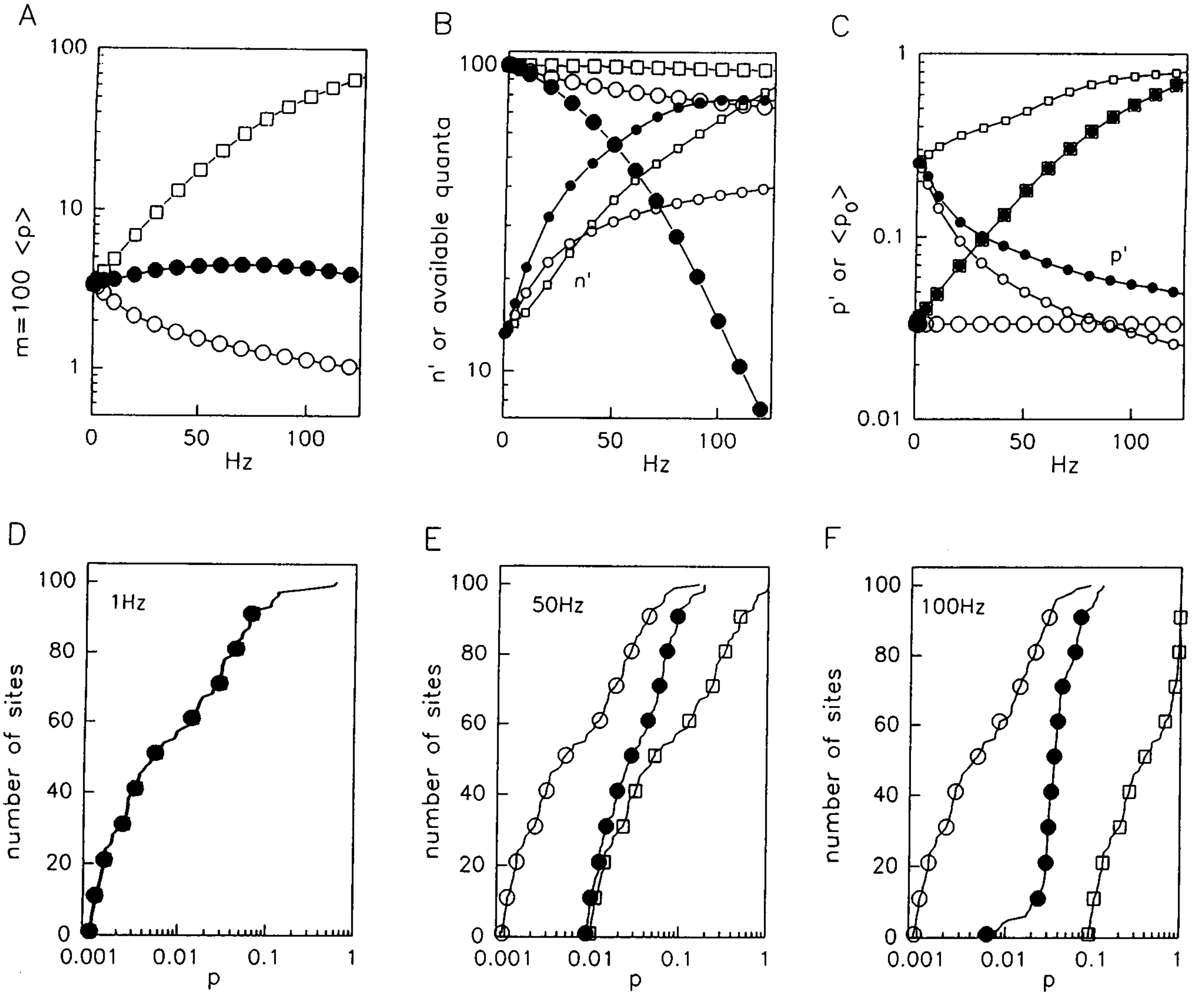

FIGURE 5 Theoretical equilibria at various stimulus frequencies for an array of N = 100 sites with widely varying r and R_A (mean 5/s) for three scenarios: 1) constant r and R_A (*open circles*); 2) r increases exponentially with frequency, 10× at 50 Hz and 100× at 100 Hz, but R_A's are constant (*filled circles*); and 3) same as 2), but R_A's also rise with stimulus frequency so as to keep α's constant (*open squares*). The number of available quanta ($N\langle p_A\rangle$) (*large symbols in B*) falls most in scenario 2), but this is accompanied by the greatest rise in n'; n' (*small symbols in B*) always rises with stimulus frequency, although p' (*small symbols in C*) may either rise (scenario 3) or fall (1 and 2). (*D*, *E*, and *F*) Cumulative distributions of p (= $p_o p_A$) at 1 Hz, 50 Hz, and 100 Hz, respectively. Distributions always narrow at high stimulation frequency, either because depletion is more at high-p_o sites and less at low-p_o sites ($p_o p_A$'s become more uniform; scenario 1), or because no p can be more than 1.0 (potentiation raises initially low p_o's more than it raises initially high p_o's; scenario 3), or for both reasons (scenario 2). In *D–F* the symbols are at every tenth point.

p_A's (1 and 2) and/or, even if refill accelerates with stimulus frequency (3), because initially high p_o's, in contrast to low p_o's, have little room to grow, being limited to the maximum of p = 1.

Of course, the extent to which n' grows with stimulation frequency depends upon the distribution and absolute magnitudes of p_o's and α's. The only scenario in which n' does not grow with stimulation frequency that I have been able to find (not illustrated) is one where α's are related to p_o's in such a way that low-p_o sites deplete as much as high-p_o sites and p_o's either do not grow with stimulation frequency or are all so low that with facilitation none approaches unity.

Simulations for equilibrium: sampling error

Using Monte Carlo simulations with up to 160,000 stimuli in each sequence, and arrays of 40 sites with widely varying p_o's, it was verified that means, variances, and covariances did not differ significantly from values obtained by using Eq. 3 and summations. Depending on α's (R_A's and stimulus frequencies), covariances between successive outputs were never more negative than −10% of the variances. Using small groups of k outputs (with subsequent averaging) to determine sampling error showed the ratio covariance/variance to be consistently biased upward by $1/k$. The standard deviation of these ratios was close to $k^{-0.5}$ (for k 〉

100, more for lower k), i.e., ~0.05 for groups of 400 outputs. Thus, one cannot envisage a statistically significant value for equilibrium covariance/variance with fewer than four groups of 400 stimuli.

Values of p' ($\equiv 1 - \mathrm{var}(m)/\langle m \rangle$) determined from values of $\langle m \rangle$ and var(m) within small groups and subsequently averaged were unbiased, but small group estimates of n' ($\langle m \rangle / p'$) were systematically biased upward, particularly when true p' was less than about 0.2; this bias became less than 5% for $p' \rangle 0.3$ and groups of 200 outputs or more. Notably, the sampling errors of p' and n' at low p' are rather large; the scatter shown in Figs. 1 *B* and 2 *B* (1000 samples at each point) was typical.

Spontaneous loss of quanta from sites

For completeness, one must consider that filled sites might at any moment become unfilled, by internal loss and/or spontaneous release, both loss and refill being continuous processes. An exact mathematical result can be obtained by dividing the time interval between stimuli (T) into a succession of small δt's and taking limits. Assume rate constants for internal loss, R_L, spontaneous release, R_o, and replenishment, R_A. In each δt loss is $(R_L + R_o)\mathrm{n}\delta t$, and refill is $R_A(n - \mathrm{n})\delta t$, where n is the number of filled sites at any moment. At the limit one obtains the differential equation $\mathrm{dn}/\mathrm{d}t = R_A n - (R_A + R_L + R_o)\mathrm{n}$. This is a standard pool model with entry and exit; n changes exponentially with rate constant $\tau = 1/(R_A + R_L + R_o)$, asymptoting to $n' = nR_A/(R_A + R_L + R_o)$. Previously α was $1 - \exp(-TR_A)$; it now becomes $1 - \exp(-T/\tau)$, and β becomes $\exp(-T/\tau)$, which is less than previously. Equations 2 and 3 are modified only by the new definition of α and β and by replacing n by n' in equations for $E(\mathrm{n_i})$ or $E(\mathrm{n_{i+1}})$ in which n appears. The net effect of loss is twofold: 1) maximum $p_A = R_A\tau$, which is <1, and 2) reduction of covariances.

Relation between p_o and stimulus-evoked release rate(s)

It has already been pointed out that p_o's must be constrained to less than unity in modeling how an array of synaptic sites may behave if p_o's are to be modulated, e.g., hypothesizing that p_o is increased by raising external [Ca^{2+}], or with facilitation. The problem of how to do this disappears if one considers p_o to result from a succession of small probabilities within a total release time t. To be precise, let us suppose that these probabilities are $r_1\delta t, r_2\delta t, \cdots, r_j\delta t, \ldots,$ etc. in succession after a stimulus. Here r_1, r_2, etc. are release rates, which can have any positive value, provided δt is made sufficiently small. Assume further that t is so brief that the refill possibility is negligible, i.e., no more than one quantum may be released. Then the chance of a quantum being released (if available) is 1 − prob(no release). The chance of a quantum from a filled site not being released in each time period i is $Q_i = \exp(-r_i\delta t)$; the chance of it not being released in the whole period t is the product of all Q_i's. Hence,

$$p_o = 1 - \exp(-\delta t(r_1 + r_2 + r_3 \cdots)) = 1 - \exp(-r_o t),$$

where r_o is the average r_i over time t. These $r_o t$'s correspond to the r's already used in modeling how outputs may change with "global" facilitation (Figs. 2, 3, and 5).

Release asynchrony

There have now been many experimental observations of the time course of release; although the major portion occurs in a time window of less than a millisecond (e.g. Katz and Miledi, 1965; Bain and Quastel, 1992a), the situation is complicated by a tail of raised frequency of "miniature" quantal events that decays with a time constant on the order of 100 ms (e.g., Hubbard, 1963; Bain and Quastel, 1992b), at least some of which may or may not—the decision is arbitrary—be included in the evoked synaptic signal.

No information currently exists on the extent to which observed dispersion may represent variation between rather than at sites. Nevertheless if within-site time dispersion of high release probability exists at all, stochastic replenishment implies that a single site sometimes releases more than one quantum after a stimulus, because there is some chance

TABLE 2 Increases in apparent quanta/site (n_a) at a single site with one quantum, and decreases of apparent p (p_a) produced by taking into account partial replenishment during release period

	$T/\tau = 0.5$		$T/\tau = 1.0$	
	$\Delta p_a/\Delta n_a = -0.95 + 0.20\mathrm{p_o}$		$\Delta p_a/\Delta n_a = -0.96 + 0.32p_o$	
p_o	p	Δn_a^*	p	Δn_a^*
0.125	0.105	0.8	0.117	0.8
0.25	0.180	1.1	0.218	0.9
0.5	0.282	1.6	0.387	1.2
0.75	0.348	2.6	0.522	1.7
0.875	0.373	3.6	0.580	2.3
0.95	0.385	4.7	0.612	2.9
0.99	0.392	6.5	0.628	4.0

	$T/\tau = 2.0$		$T/\tau = 4.0$	
	$\Delta p_a/\Delta n_a = -0.98 + 0.45p_o$		$\Delta p_a/\Delta n_a = -0.99 + 0.55p_o$	
p_o	p	Δn_a^*	p	Δn_a^*
0.125	0.123	0.7	0.125	0.7
0.25	0.241	0.8	0.249	0.8
0.5	0.464	1.0	0.495	0.9
0.75	0.671	1.3	0.740	1.2
0.875	0.770	1.7	0.861	1.5
0.95	0.827	2.2	0.933	1.9
0.99	0.857	2.9	0.972	2.5

Listed Δn_a^* is the percent change in apparent n (from unity) for $t_{\mathrm{eff}}/\tau = 0.01$; for other t_{eff}/τ Δn_a is proportional to t_{eff}/τ. In each case the first column on the left gives p_o, and listed p is $p_o p_A$. Effective refill between stimuli is always $1 - \exp(-T/\tau)$. Changes in apparent p (Δp_a) are proportional to Δn_a and dependent upon p_o according to the formula.

of refill while release probability is still high. The effects of this on means and variances were calculated using the general scheme above, with a succession of small p_o's in small δt's. The results are shown in Table 2, which is explained further in the Appendix. The general result is that each site indeed behaves on average as though it had more than one quantum available for release (apparent n, $n_a > 1$), but apparent p (p_a) is decreased; mean output is increased less than variance. By and large, effects are small if most release occurs within a time that is on the order of 1% or less of the replenishment time constant. In the rest of this section it will be assumed that instantaneous release is a valid approximation for release by stimuli.

Modification of statistical measures by quantal amplitude and stochastic channel closing

Rarely can released quanta be counted directly in an experiment. Instead one measures signals that represent responses to individual or summed quanta. Assuming linear summation, if quanta all give rise to a response of constant amplitude (or area, if time integrals are measured) h, the mean response is scaled by h, the variance and covariances by h^2, and the third moment by h^3. Otherwise, the scaling depends on whether nonuniformity in h occurs at every release site, or whether h varies between release sites, or both; models currently employed in the analysis of CNS synaptic signals differ in this respect (Redman, 1990; Walmsley, 1993; Jack et al., 1994).

(a) Quantal responses are constant at each release site but vary between sites

In this case, mean, variance, covariances, and third moment are scaled at each site by its h, h^2, h^2, and h^3, and the mean, variance, etc. for the array are obtained by addition, i.e., the only general formulae are $E(S) = \Sigma\,(hp)$, $\mathrm{var}(S) = \Sigma\,(h^2(p - p^2))$, etc., where S denotes either signal height or area. Notably, var(S) is less than if h varies at each release site. For example, suppose there are three release sites with $h = 1$, $h = 2$, and $h = 3$, respectively. Then a "success" at every site (three quanta) always has $S = 6$, but if quanta may have $h = 1$, 2, or 3 at each site, S can vary between 3 and 9.

(b) Quantal responses vary at every release site (and not between)

In this case moments can be calculated directly. The moments about 0 at each release site, μ_1', μ_2', and μ_3', are each p times the respective moments of h about 0, i.e., $p\langle h\rangle$, $p(\langle h\rangle^2 + \mathrm{var}(h))$, and $p(\langle h\rangle^3 + 3\langle h\rangle\mathrm{var}(h) + H3)$, respectively, where $H3$ is the third moment of h about its mean. Therefore, for each release site,

$$\begin{aligned} E(S) &= \langle S\rangle = \mu_1' = \langle h\rangle \mathrm{P} \\ \mathrm{var}(S) &= \mu_2' - (\mu_1')^2 = \langle h\rangle^2 p(1 + cv_\mathrm{h}^2 - p) \\ S3 &= \mu_3' - 3(\mu_1')(\mu_2') + 2(\mu_1')^3 \\ &= \langle h\rangle^3 p[(1-p)(1-2p) + 3p(1-p)cv_\mathrm{h}^2 + h3')]. \end{aligned} \tag{4}$$

Here $S3$ is the third moment of the signal; cv_h is the coefficient of variation of h and $h3'$ is $H3/\langle h\rangle^3$.

Summing over N release sites gives

$$\begin{aligned} E(S) &= \langle S\rangle = N\langle p\rangle\langle h\rangle = \langle m\rangle\langle h\rangle \\ \mathrm{var}(S) &= \langle S\rangle\langle h\rangle(1 - p' + cv_\mathrm{h}^2) \\ &= \langle h\rangle^2(\mathrm{var}(m) + \langle m\rangle cv_\mathrm{h}^2) \end{aligned} \tag{5}$$

and

$$S3 = \langle h\rangle^3[M3 + \langle m\rangle(h3' + (1 - p')cv_\mathrm{h}^2)],$$

where p' is, as before, $\langle p\rangle$ $(1 + cv_\mathrm{p}^2)$ and $\langle m\rangle$ has been equated with $E(m)$. The first expression for var(S) gives p' and hence var(m) if cv_h as well as $\langle h\rangle$ can be inferred from "miniatures."

The covariance between successive responses to stimuli is, of course, simply scaled by $\langle h\rangle^2$, assuming that there is no covariance of quantal amplitude from one stimulus to the next.

It is notable that the coefficient of variation of the signal, cv_S, is given by

$$cv_\mathrm{S}^2 \equiv \mathrm{var}(S)/\langle S\rangle^2 = (1 - p' + cv_\mathrm{h}^2)/\langle m\rangle.$$

For a Poisson distribution, with $p' = 0$, $\langle m\rangle$ is given by $\langle S\rangle^2/\mathrm{var}(S)$ if $cv_\mathrm{h}^2 = 0$. Because p' is usually less than 0.5, and cv_h^2 is unlikely to be more than 1, $\langle S\rangle^2/\mathrm{var}(S)$ generally gives an approximation of $\langle m\rangle$ that is accurate within a factor of 2 or so.

(c) Quantal responses vary both between sites and at sites

In this case one sums as in (a), taking into account the variance, etc., of h at each site using Eq. 4:

$$E(S) = \sum(\langle h\rangle p); \qquad \mathrm{var}(S) = \sum(\langle h\rangle^2 p(1 + cv_\mathrm{h}^2 - p))$$

$$S3 = \sum(\langle h\rangle^3 p[(1-p)(1-2p) + 3p(1-p)cv_\mathrm{h}^2$$

Contribution to the variance of stochastic channel closing: estimation of channel amplitude

If signals are voltage signals with the decay rate dominated by the cell input impedance, the coefficient of variation of signal areas (time integrals) is the same as that of signal heights, but if currents are measured it is greater, because of the contribution to variance of stochastic closing of the channels underlying the quantal responses. For exponentially distributed channel durations I calculate (see Appendix 3) that cv_S^2 for the signal area is higher than for height by just $1/\langle n_c\rangle$ whatever the signal-to-signal distribution of n_c, the number of open channels in each. An important assumption here is that all quantal responses and all channel openings occur close enough in time that net signal height represents all channels that open. If not, one could obtain $\langle n_c\rangle$ and hence channel amplitude, h_c, from means and variances of signal heights and subsequent areas somewhere in the decay phase, beyond which there is no new channel opening. Of course, to estimate h_c in this way, artificially aligned miniatures would serve as well as responses to stimuli, and in either case the noise component of variances would have to be subtracted.

The area product

In practice the determination of means and variances (and covariances) from experimental data is not simple. It generally involves finding a baseline for each signal and a decision whether to measure maximum height, the height at the average maximum (the latter is unbiased by noise but more biased by time dispersion of release), or signal area, which is relatively sensitive to any error in baseline.

A way to avoid such problems and to extract some added information from the data is to make use of the covariances of point values with signal sums. For want of a better term, I call these the "area product" (A). This function can be determined by accumulating from the record for each stimulus (y_1, y_2, y_3, . . . etc.), including prestimulus values, the product of each point value y_j with the sum of all values in the record (S), while also obtaining mean y_j's and mean S. Designating as A_j the covariance at point j, it turns out that the sum of A_j's for all points in the record is exactly the same as the variance of S, as calculated simply from values of S for every record (see Appendix, 4 i):

$$\sum A_j = \mathrm{var}(S) \tag{6a}$$

where the summations (for S and ΣA_j) are over all j, to where the signals have decayed to a small fraction of maximum. Moreover, if certain conditions are met,

$$A_j = E(y_j)\,\mathrm{var}(S)/E(S). \tag{6b}$$

To derive Eq. 6b, consider the signals obtained from a single release site that releases quanta with a time sequence of probabilities giving rise to a release time course $p(t)$ with total probability p ($= \Sigma\, p(t)$'s). Each quantal response has a time course $h(t)$, the mean signal being the convolution of $h(t)$ and $p(t)$. For simplicity let us choose a time base so that the time integral of the response, $\Sigma\, h(t)$, is equal to unity. We have uncorrelated records. In any single record there is either no quantum or one; signal area (S) is 0 or 1. Therefore, for each record the cross-products of point values with S are the same as the original record (Fig. 6 A). For all of the records, at point j, this product has the expected value $E(Sy_j) = E(y_j)$. The expected value of mean area, $E(S)$, is p. The covariance of the point value with the sum of the signal (S), i.e., $\langle(y_j - \langle y_j\rangle)(S - \langle S\rangle)\rangle$, is

$$\begin{aligned}\text{Area product for 1 site at point } j = A_j\\ &= \mathrm{cov}(y_j, S) = E(Sy_j) - E(S)E(y_j) = E(y_j) - pE(y_j)\\ &= (1-p)E(y_j)\end{aligned}$$

Because with quantal area = 1, $\mathrm{var}(S)/E(S) = (1 - p)$, this result corresponds to Eq. 6b.

If the quantal response has area h, S is either h or 0, $E(Sy_j) = hE(y_j)$ and $E(S)$ is hp; $\mathrm{cov}(y_j, S) = h(1 - p)E(y_j)$, and Eq. 6b is again correct. In Fig. 6 A the terminology is slightly different; here the quantal responses decay exponentially and have height h and area $h\tau$, to contrast the result with what occurs with single channels, with exponentially distributed lifetimes (Fig. 6 B), having the same height and mean area as the quanta in Fig. 6 A.

Now it is important to note that the area product at each point is a covariance and therefore is additive for independent stochastic processes, and that because Eq. 6a is a numerical identity, scaling of the area product by quantal height and variance is just the same as that of variance of signal sums.

Summing over N sites with different p's, one finds that there are two essential provisos for the area product function $A(t)$ to recapitulate the time course of $\langle y(t)\rangle$ (Eq. 6b), namely 1) every quantal response has the same time course, $h(t)/h$, and 2) quanta of different amplitudes have the same $p(t)/p$. Thus the time partitioning of var(S) provided by $A(t)$ can indicate whether these provisos are not met:

1. Release sites producing quanta of different amplitudes have differing $p(t)/p$. Example: Release at large quanta sites is delayed until the signal from small quanta has decayed—early values of $A_j/\langle y_j\rangle$ pertain to small quanta and late values to large quanta.

2. If stimulation causes the appearance of quantal responses differing in $h(t)/h$, the more prolonged contribute more to late values of A_j and the ratio $A_j/\langle y_j\rangle$ is eventually that expected for the most prolonged alone. Example: One set of quantal responses is "filtered" by electrotonic conduction and therefore is prolonged, whereas others are not (see e.g., Jack et al., 1994).

3. If there is postsynaptic nonlinear summation of guantal responses, the height of quanta is in effect reduced as quanta are superimposed: $A_j/\langle y_j\rangle$ characteristically dips when $\langle y_j\rangle$ is high.

4. *Always*, with voltage clamp, because quantal responses, each a composite of currents through a number of channels with stochastically varying lifetimes, never have absolutely identical time courses.

The last breach is of particular interest as a common complication, and because it leads to a simple method of determining unit channel amplitude, provided other complications can be ruled out, the number of channels per quantal response is fairly low, and recording noise is not overwhelming.

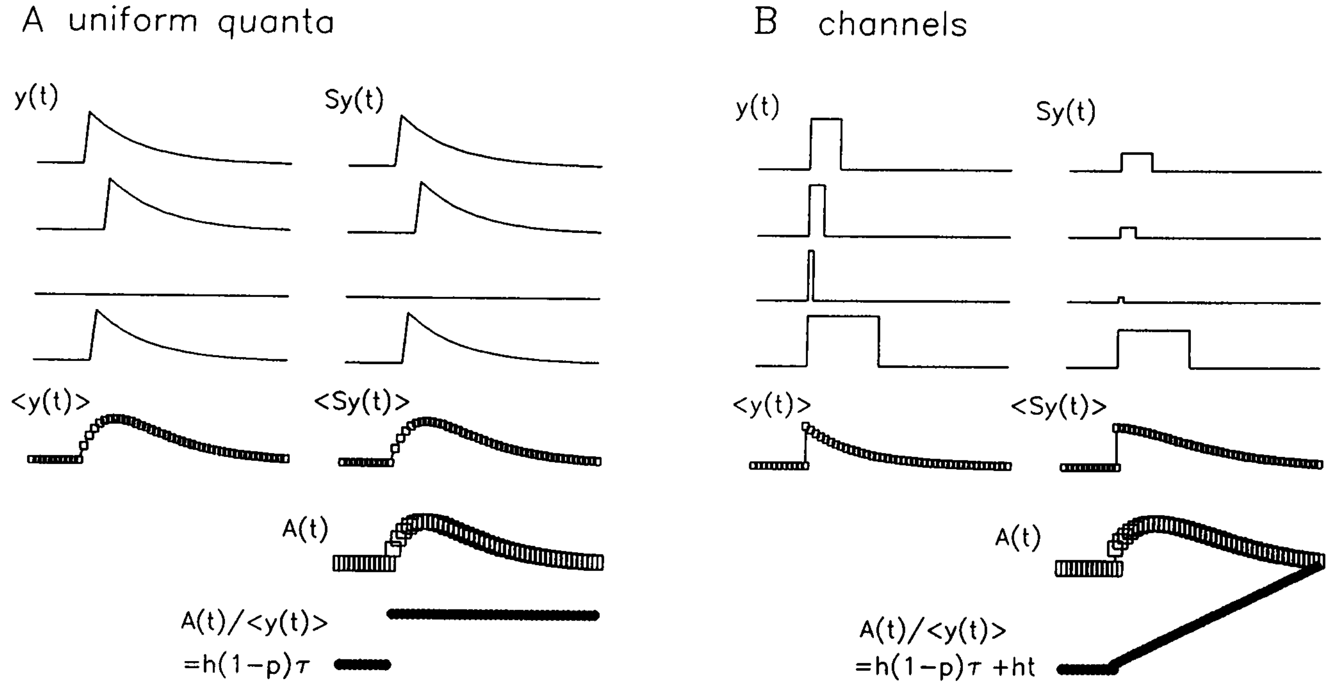

FIGURE 6 The theoretical basis for the area product. In *A* a single site on stimulation may or may not release, with a time-distributed probability *p*, a quantum producing an exponentially decaying response with unit area. Four samples of individual responses are shown above and to the left, and the corresponding cross-products with area, which are the same. Theoretical averages for a large number of such records are shown below, with the corresponding *A*(*t*) and the ratio $A(t)/\langle y(t)\rangle$. The latter is $(1 - p)$ multiplied by the quantal area, height (*h*) multiplied by time constant (τ). In *B* the unit response is the opening of a single channel of height *h*, with exponentially distributed lifetimes with mean τ. Brief channels give a brief small *Sy*(*t*), whereas prolonged channels give prolonged large *Sy*(*t*); the ratio of the area product (*A*(*t*)) to $\langle y(t)\rangle$ rises linearly with slope *h* (see Appendix, 4 ii). In the plots shown the scales of *A*(*t*) and $A(t)/\langle y(t)\rangle$ have been chosen for convenience.

In Fig. 6 *B*, a patch with a single channel is envisaged; each channel has an *Sy*(*t*) that is a square pulse with height proportional to its duration; durations are exponentially distributed. For channels with mean duration τ_c and amplitude h_c, opening at the same time with probability *p*, summing all products of probability and outcome (see Appendix, 4 ii) gives

$$A(t) = \langle y(t)\rangle(h_c\tau_c(1 - p) + h_ct), \qquad (6c)$$

where $\langle y(t)\rangle$ is, of course, $h_cp\exp(-t/\tau_c)$.

As it turns out, the linear growth of $A(t)/\langle y(t)\rangle$ versus *t* with slope h_c remains if quantal responses reflect groups of channels opening; the expected value when the signal begins is the same as if quantal responses were uniform in time course (see Appendix 4 ii). Moreover, taking as unit time the sampling interval (i.e., simply adding point values to make sums), $A_j/\langle y_j\rangle$ grows linearly with *j*, with slope h_c.

If channels do not all open simultaneously, the theoretical value of *A*(*t*) is more complicated. However, calculations (and simulations; Fig. 7) show that the linear growth of $A_j/\langle y_j\rangle$ with slope h_c remains, at least after the peak of the signal, provided most openings occur before most closings. If channels flicker between open and closed states (these closings do not count in the previous sense), but net open times remain exponentially distributed, then the h_c one obtains is the true h_c multiplied by the fraction of time that the channels are open. This is equally true for h_c found from the coefficients of variation of signal height and area (see above), which rests on the same assumptions about channel behavior. Moreover, in both cases, the "extra" variance disappears with sufficient electrotonic filtering. To obtain h_c in practice, supposing that all but one kind of channel have been eliminated pharmacologically, and responses are from a set of synapses with much the same *p*(*t*)/*p*, one would obtain parameters for the least-squares best fit to baseline normalized $A_j/\langle y_j\rangle = a + bj$, for *j* past the peak of $\langle y_j\rangle$, including only *j*'s with well-defined $\langle y_j\rangle > 0$, and weighting by $\langle y_j\rangle^3$ (see Appendix, 5), with *b* being the putative value of h_c. As with the other method for determining h_c, this is equally applicable to artificially aligned spontaneous miniatures, or, indeed, evoked signals grouped according to peak amplitude.

Noise

As a stochastic process not time-locked to the stimulus, recording noise contributes a positive value to *A*(*t*) that is the same at all times, including prestimulus; for any simple low-pass filter, the expected value turns out to be the noise variance of the unfiltered records. Noise also adds somewhat to the noisiness of the area product; but this effect is small if responses can be seen at all. In the simulation in

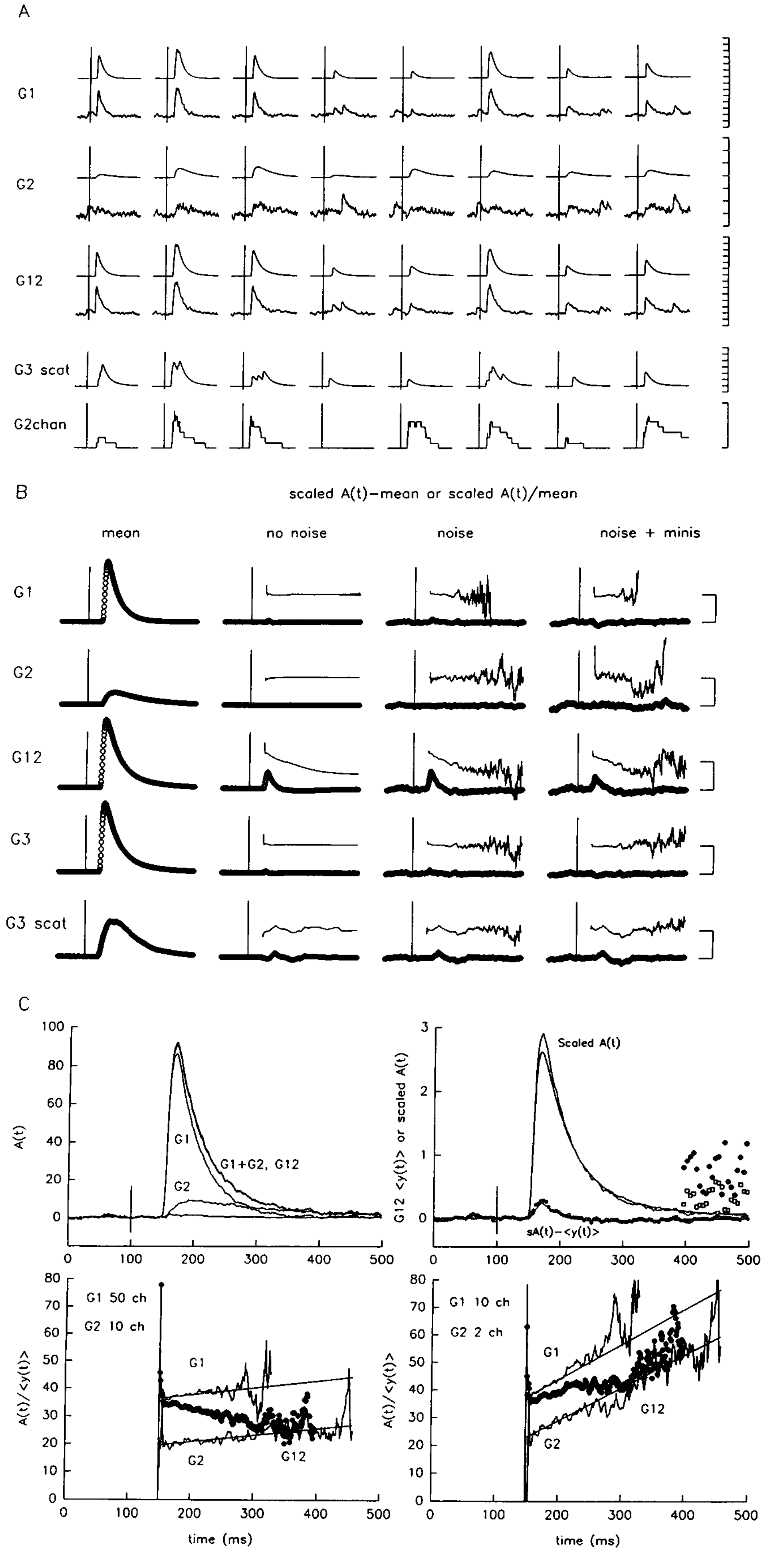

FIGURE 7 Realization of area products for Monte Carlo simulated data for arrays of 10 sites of each type with divers p_o's, and refill between stimuli so as to give $\langle p \rangle = 0.25$, $\langle m \rangle = 2.5$, and $p' = 0.33$. In *A* are samples of simulated records (note amplitude scales), and in *B* the means and differences from the means of scaled area products (see text), for stimulation of synapses with large brief quanta (G1 height 1 unit, decay τ 40 ms/points) or smaller prolonged quanta (G2 height 0.17 units decay τ 120) or stimulation of both together (G12). G3 refers to synapses where each quantal response is the sum of a G1 and G2 quantal response, and the samples shown in *A* (G3-scat) pertain to highly time-dispersed release. The samples for G1, G2, and G12 in *A* are both without and with simulated Gaussian noise (rms 0.5 units, enough to make unit G2 quantal responses invisible), plus spontaneous miniatures of both G1 and G2 types, and for illustration have been low-pass filtered to increase the visibility of the responses. The graphs in *C* illustrate the additive properties of the area product (first graph), nonequivalence of time course of $A(t)$ and $\langle y(t) \rangle$ (second graph) for a mix of quantal types (G12), and linear rise in $A(t)/\langle y(t) \rangle$ versus t for G1 and G2 alone but not for the mix (*filled circles*). These graphs pertain to simulations with added noise in which G1 and G2 responses have on average 50 and 10 channel openings, or 10 and 2, respectively (last graph only; latter are also shown without noise as G2-chan in *A*). The straight lines in the two last graphs in *C* are theoretical, corresponding to the expectation for the respective channel amplitudes.

Fig. 7 (see below), so much noise was added that the small quantal responses cannot be seen, but 200 records with 500 quanta in all yield a very "clean" $A(t)$.

Spontaneous miniatures may constitute an important source of "noise." This, too, can be eliminated by subtracting from all values of $A(t)$ the average prestimulus value. However, if the total area of miniatures in records is much more than the total area of signals, the resulting point-to-point noisiness of $A(t)$ may limit the usefulness of the area product to determination of noise-unbiased var(S) by summing A_j's after baseline correction.

Monte Carlo simulations

Fig. 7 illustrates how well area products conform to expectation and extract otherwise inaccessible information from simulated data (samples in Fig. 7 *A*). From series of 200 synaptic signals with $\langle m \rangle = 2.5$, and uniform quanta, area products had virtually the same time course as averages (Fig. 7 *B*), either with relatively large brief quantal responses (G1) or small prolonged quanta (G2), or quanta each consisting of the sum of a G1 type and G2 type quantum (G3). To facilitate visual comparison in Fig. 7 *B*, the area products, after subtracting prestimulus values, were scaled to have the same area as the means; the differences of scaled area product, $sA(t)$, from means ($\langle y(t) \rangle$) are plotted to the right of the means in Fig. 7 *B* for three scenarios: no recording noise, Gaussian noise sufficient to obscure quanta of the G2 type, and the same noise plus random miniatures of both G1 and G2 types at an average of one each per record. In contrast, for synaptic signals corresponding to G1- and G2-type synapses being stimulated simultaneously (G12), the differences are substantial. Correspondingly, plots of ratios of $A(t)/\langle y(t) \rangle$ versus time (*thin lines*) are flat (but noisy when $\langle y(t) \rangle$ is low), except for the G12 situation, where this ratio declines as smaller, more prolonged G2 quanta become the predominant components of both $A(t)$ and $\langle y(t) \rangle$. In the G3 scat simulation, highly time-dispersed G3 quantal release (see samples in Fig. 7 *A*), the difference of $sA(t)$ from $\langle y(t) \rangle$ wobbles, probably because about 500 quanta was insufficient for the actual release dispersion to closely approximate $p(t)/p$ late in the signals.

For Fig. 7 *C* the simulation was continued to 1000 records, for a situation in which G1 quanta contained on average 50 channels and those of the G2 type 10 channels, so that h_c (but not τ_c) was the same for both types. Simulated recording noise was included in these records, but not minis. The first graph illustrates the additive properties of $A(t)$ for the mix (G12); $A(t)$ is indistinguishable from the sum of $A(t)$'s for G1 and G2 alone, the difference (*line near zero*) being negligible at all times. The second graph shows $\langle y(t) \rangle$ for the mix, and its scaled $A(t)$; the differences (*small circles*) are essentially identical to those expected (*nearby line*) for an $A(t)$ equal to the sum of $A(t)$'s for G1 and G2 alone. The filled circles on the right show SDs of $sA(t)$ and $y(t)$, respectively (each ×100), illustrating that the intrinsic noisiness of the area product is not much more than the noisiness of the mean.

The third graph in Fig. 7 *C* shows how for G1 or G2 alone the ratio $A(t)/\langle y(t) \rangle$, past the peak of the signal, indeed grows linearly with a slope of h_c per point, whereas for the mix (G12) the ratio gradually falls from an appropriate value for G1 alone (at the start where G2 responses are negligible) to the value for G2 alone. This is also seen in the last graph, where the model was modified by ascribing an average of 10 channels to G1 quanta and an average of 2 channels to G2 quanta (also see the bottom sample in Fig. 7 *A*). The noisiness of the plotted ratios (and initial high values) comes from including points with very low $\langle y(t) \rangle$'s that would not be included in finding h_c by least-squares fitting.

However, in the last graph in Fig. 7 *C*, an intrinsic ambiguity of the area product is exemplified in that with the output mix (G12 *filled circles*), the plot of $A(t)/\langle y(t) \rangle$ versus time could here be mistaken for that produced by a single set of quanta, i.e., flat until too noisy for the late rise to be ascribed to anything but noise. Using subgroups of signals selected by amplitude would resolve this ambiguity.

Dealing with drift and finding sequential signal covariance

Another problem in analysis is how to compensate for any drift (nonstationarity) in the signal, i.e., m's and/or h's trending up or down. This can add substantially to variances (and reverse covariances); being able to compensate adds to the variety of usable stimulation paradigms (e.g., Elmqvist and Quastel, 1965). Assuming stimulation at a constant frequency, one way to exclude effects of such drift is by determining variance (and covariance) within small groups of sequential records. The smallest possible group is 2; determination of the area product now reduces to taking for each record the point-to-point difference of this record ($y_{i,j}$'s) from another nearby ($y_{i+k,j}$'s) to obtain a new series of numbers (z_j's) and accumulating products ($z_j \, \Sigma \, z$). Corresponding to Eq. 6, the expected mean of these turns out to be

$$E(z_j \sum z)/2 = A'_j = E(y_j)(\mathrm{var}(S) - \mathrm{cov}(S_i, S_{i+k}))/E(S) \quad (7a)$$

$$\sum A'_j = \mathrm{var}(S) - \mathrm{cov}(S_i, S_{i+k}). \quad (7b)$$

Because $\mathrm{cov}(S_i, S_{i+k})$ is negligible for $k \geq 2$, using differences between records both one apart and two or more stimuli apart gives both var(S) and $\mathrm{cov}(S_i, S_{i+1})$. With simulations, values of equilibrium variance determined in this way were found to be unbiased, but had a sampling error increased by about 50%; using two such differences for every output reduced to no more than 10% the increase in sampling error. Obtaining covariances by also using differences between sequential outputs gave sampling errors for covariance/variance 25% higher than when determined in the usual way, and no bias.

An alternative that essentially eliminates even rapid drift effects is to take the sum of point-to-point differences from values k stimuli before and k stimuli after, i.e., $z_j = 2y_{i,j} - y_{i+k,j} - y_{i-k,j}$. Then,

$$\Sigma_j\, E(z_j\, \Sigma\, z) = 6\ \mathrm{var}(S) - 8\ \mathrm{cov}(S_i, S_{i+k}) + 2\ \mathrm{cov}(S_i, S_{i+2k})$$

Point-to-point variance and the third moment of S

The area product should not be confused with the point-to-point variance of signals, which is quite different. At a single site, with one quantum to release, one has generally

$$\mathrm{var}(y(t)) = p(t) * [\langle h(t)\rangle^2(1 + cv^2_{h(t)})] - (p(t) * \langle h(t)\rangle)^2$$

where $*$ denotes convolution. In the case of a Poisson distribution of outputs, which provides an excellent approximation when $p' \ll 1$, and multiple sites, this equation reduces to

$$\mathrm{var}(y(t)) = m(t) * [\langle h(t)\rangle^2(1 + cv^2_{h(t)})]$$

where m(t) is the time course of quanta appearing. If quantal responses consist of currents through not too many channels, $cv^2_{h(t)}$ can become substantially different from $cv^2_{h(0)}$ because of the "extra" variance introduced by stochastic channel closing. For channels of uniform amplitude, opening simultaneously and closing randomly, one has for point-to-point variance of quantal responses,

$$\mathrm{var}(h(t))/\langle h(t)\rangle = h_c + \langle h(t)\rangle(cv^2_{h(0)} - h_c/h(0))$$

which, of course, provides yet another method for obtaining h_c from artificially aligned miniatures. The terms with h_c disappear if signals are overfiltered (e.g., electrotonically, or if one is recording voltage) with a time constant $>\tau_c$.

For a Poisson distribution, the third moment of $y(t)$ is given by $m(t) * [\langle h(t)\rangle^3(1 + x)]$, where x is the sum of terms that disappear if quanta are uniform in height and time course.

For the third moment of S ($S3$), the analog of Eqs. 6a and 6b is obtained by taking the mean product $B_j = \langle(y_j - \langle y_j\rangle)(S - \langle S\rangle)^2\rangle$. The sum of B_j's is numerically identical to $S3$, and $B(t)$ has the same time course as $\langle y(t)\rangle$, subject to the same provisos as for $A(t)$. For quanta that vary in time course because of stochastic closing of channels, $B_j/\langle y_j\rangle - 2A_j/\langle y_j\rangle$ grows linearly with j^2, with slope h_c^2.

"Spontaneous" release

In the original quantal analysis of synaptic signals, at the neuromuscular junction, it was shown that the quantal components of the stimulation-evoked synaptic signal correspond (in amplitude, time course, and sensitivity to postsynaptically acting drugs) to the "miniature" signals that represent spontaneous release of quanta of neurotransmitter (del Castillo and Katz, 1954a, 1956). There are now many reported examples of miniatures at diverse synapses.

The frequency of "spontaneous" miniature signals can be increased in a variety of ways, but from the statistical point of view it does not matter whether release is truly spontaneous or is evoked by a steady stimulus such as nerve terminal depolarization or raised osmotic pressure. In either case, miniatures occur apparently randomly. Vere-Jones (1966) has shown that if there are a limited number N of release sites and at each site quantal discharge is followed by a waiting time (presumably stochastic and exponentially distributed) before release can again occur, outputs will be underdistributed relative to a Poisson. This is manifested in three ways: 1) the variance is less than the mean for numbers of miniatures in nonoverlapping time periods; 2) there is a small negative covariance between such numbers; and 3) the rate of occurrence of miniatures is transiently diminished after each miniature.

Variance, means, and covariances of outputs in nonoverlapping time periods

These are derived from Eq. 3 by envisaging a succession of small p_o's (and α's) in small δt's, adding outputs for a given time period T, and obtaining variance and covariances of these outputs by adding variances and appropriate (negative) covariances. By taking limits, I obtain the following for outputs (o) from a single release site with rate of release from a filled site, R_o, rate of refilling of an empty site, R_A, and rate of internal loss from a filled site, R_L. The net release rate, R, is equal to $R_oR_A\tau$, where τ, the time constant, is $1/(R_o + R_L + R_A)$ and $R_A\tau$ is the expected number of filled sites at any moment (see also above):

$$\mathrm{mean} = E(o) = RT$$

$$\mathrm{var}(o) = RT\{1 - 2R\tau[1 - (1 - e^{-W})/W]\}$$

$$\mathrm{cov}(o_i, o_{i+1}) = -[R\tau(1 - e^{-W})]^2 \tag{8}$$

$$\mathrm{cov}(o_i, o_{i+k}) = \mathrm{cov}(o_i, o_{i+1})e^{-(k-1)W}$$

where W denotes T/τ. The formula for variance differs from Vere-Jones (1966), which has a misprint.

Ignoring the hypothetical R_L, the ratio of variance to mean is close to unity if $R_o \ll R_A$ or $R_o \gg R_A$, or if $T \ll \tau$; otherwise the ratio progressively declines as T is increased, to $1 - 2R\tau$, which has a minimum of 0.5 at $R_o = R_A$. The covariance between successive outputs, $\mathrm{cov}(o_i, o_{i+1})$, has a (negative) maximum relative to variance when T is somewhat greater than τ; the ratio of covariance/variance is most negative when $R_o = R_A$, at which it is only -0.133.

For N sites with different R_o, etc., one adds means, variances, and covariances to obtain the behavior of the array. I find that if R_o varies widely and R_A is similar at the N sites or varies randomly between sites, minimum variance/mean is about 0.6 and is found at T about $10/\langle R_A\rangle$ and with $\langle R\rangle T$ on the order of one per site. If at some sites R_o is more than R_A, a T can usually be found at which covariance/variance is about -0.1 (i.e., is potentially detectable experimentally with long data runs), in a situation in which miniatures represent release from a few sites at rates that produce major depletion.

Net release rates after each miniature: the autocorrelogram of miniatures

Imagine miniatures from a single release site. After one appears, the next cannot occur until the site is refilled. The time constant for refilling is $\tau = 1/(R_o + R_A + R_L)$. The overall release rate, R, is $R_oR_A\tau$. It turns out that the expected rate of occurrence of miniatures, after a miniature put at time 0, is

$$R(t) = R(1 - e^{-t/\tau}). \qquad (9a)$$

This may look wrong, because the average of $R(t)$ is less than R. However, the times at which a quantum is released from a site are not typical—they are times at which a quantum is available.

To obtain the expected rate after every miniature, for N independent release sites, with release rates R_1, R_2, etc., one notes that 1) depletion occurs only at the site from which that particular miniature came—the subsequent occurrence of miniatures from other sites is uncorrelated, and 2) the chance of a miniature being from site i is R_i divided by f_m, the overall release rate or miniature frequency, which is the sum of all R's. The summation is simple if at all sites τ is much the same. At time t, the expected rate of occurrence of miniatures from all sites after a miniature from the ith site is

$$R_i(t) = f_m - R_ie^{-t/\tau} \text{ with probability } R_i/f_m.$$

Summing products of expected rate and probability,

$$\begin{aligned} f_m(t) &= \sum R_i - e^{-t/\tau} \sum (R_i)^2/f_m \\ &= f_m - e^{-t/\tau}\langle R\rangle(1 + cv_R^2) \qquad (9b) \\ &= f_m(1 - e^{-t/\tau}(1 + cv_R^2)/N), \end{aligned}$$

where cv_R is the coefficient of variation of R. Thus, if R's were the same, and N not too large, extrapolation to $t = 0$ would give N; instead it gives $N/(1 + cv_R^2)$. Except for scaling, $f_m(t)$ is the same as the autocorrelogram (i.e. autocorrelation function) of miniatures.

Simulation

In Fig. 8 spontaneous release comes from one site with R_o = 1/s and four sites with R_o = 0.1/s, all with R_A = 5/s and negligible R_L, giving a total output rate of 1.225/s (0.833/s + 4 × 0.098/s). With a total 1000-s sample time (1220 minis), the graph of f_m versus time after each mini (Fig. 8 *A*, *open circles*) shows an early drop, indicating that most release is from no more than two sites; with a very long data series (five sets of 8000 s; *filled points*) the observed function is very close to the theoretical (*line*). The ratios of variance/mean of numbers of miniatures in nonoverlapping time periods of varying duration, T, clearly become less than 1.0 at sufficiently long T, in close conformity to the theoretical curve (Fig. 8 *B*), even for the shorter recording period.

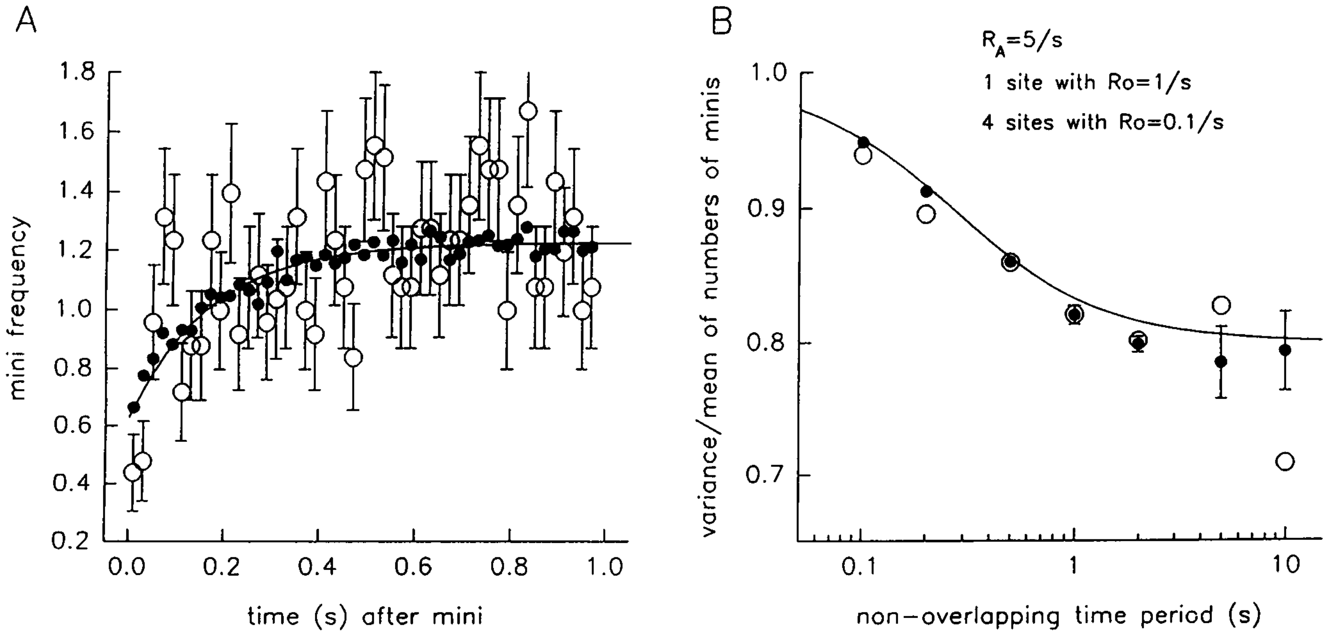

FIGURE 8 Spontaneous release. (*A*) The frequency of miniatures following a miniature (the autocorrelogram) and (*B*) variance/mean ratios of numbers of miniatures in nonoverlapping time periods, for an array with only five sites, one with R_o = 1/s and four with R_o = 0.1/s, with R_A = 5/s for both kinds of sites. The open circles represent values obtained for a 1000-s simulation period (total of 1220 miniatures), and the small filled circles are for 8000 s repeated five times (to give SE's in *B*), showing how simulation eventually fits theoretical values (*lines*).

DISCUSSION

For any series of fluctuating synaptic signals there is a mean and variance, and if the mean and variance of the quantal components are known, one can obtain the mean, $\langle m \rangle$, and variance, var(m), of numbers of quanta released, and, from these, $p' = 1 - \text{var}(m)/\langle m \rangle$ and $n' = \langle m \rangle / p'$. What is the meaning of n' and p', except as parameters that sometimes (Table 1) give a good approximation of the output distribution? In terms of the simple binomial model, in which all N independent release sites are equivalent (have the same probability of release p), n' is N, which must be invariant under all experimental conditions and no less than the maximum output obtainable under any conditions. Hence experimental results showing n' to vary with stimulation frequency and/or ambient Ca^{2+}/Mg^{2+}, at both vertebrate and invertebrate neuromuscular junctions (for references see McLachlan, 1978), negate this model. The question now arises whether n' might instead represent the number of active sites, rather than the total number. Here two distinct untenable models seem to be confounded in the literature. In the first, an active site is one with a quantum available; n' is their number and p' is the (uniform) output probability of these, here designated p_o. For this to apply, the number of available quanta would have to be constant from stimulus to stimulus, despite fluctuations in release—i.e., the model presupposes a presynaptic monitor that counts outputs and then adjusts the available quanta at every site accordingly. Apart from its implausibility, this model denies the principle of release site independence, invalidating the use of binomial statistics. The second model is a constrained special case of the compound binomial in which n' active sites all have the same $p' = p = p_o p_A$, p_A being the probability at any moment that an active site has a quantum available, the other N-n' sites being silent ($p = 0$). This model is self-contradictory, because release sites are presumed to be both identical and dissimilar simultaneously; if sites are active and can ever release a quantum, they do so with the same probability, but at the same time each site differs from others with respect to an absolute threshold for an extracellular Ca^{2+}/Mg^{2+}, or for stimulus frequency, depending on which happens to have been varied experimentally, below which it is silent. Thus the mathematical simplicity of the simple binomial model is incompatible with any simple model of presynaptic function; output distributions and fluctuation analysis based on mean and variance cannot indicate the number of active sites.

With the unconstrained compound binomial model, no such implausible partition arises. Once it is accepted that p_o and/or p_A and therefore p may vary between sites, n' is seen to be $N/(1 + cv_p^2)$ and, defining N as the total number of sites that might release a quantum at any time, any observed change in n' must by definition reflect a change in cv_p and therefore gives information as to how the distribution of p varies with, e.g., stimulation frequency or change in extracellular Ca^{2+}.

When one considers that release entails "depletion" (a temporary incapacity to release again after release of a quantum), which is also an assumption, albeit usually hidden, in the simple binomial, a stochastic repletion model (Vere-Jones, 1966) gives p_A as a function of replenishment rate, time between stimuli, and p_o, and as a result cv_p and n' can be expected to vary with the experimental situation. In particular, n' should increase with stimulus frequency, because cv_p declines automatically as sites with high p_o's develop low p_A's (Eq. 3a), unless refill rates match p_o's in just such a way to prevent this. Furthermore, an increase in n' is predicted for any global increase in p_o's (or, rather r's, with each $p_o = 1 - e^{-r}$), because the limitation of p_o to a maximum of unity at each site implies that it can grow less at initially high-p_o sites than an initially low-p_o sites, with consequent reduction of cv_p. The many examples in the literature of paired-pulse depression attributable to "depletion" (from Liley and North, 1953, to Mennerick and Zorumski, 1995) that disappears with lowered outputs in low Ca^{2+}/high Mg^{2+} are fully consistent with it being p_o's that alter with Ca^{2+} entry per stimulus.

The compound binomial model is also not without its implausible element, namely that release sites apparently differ enormously in their tendency to release. The variation between sites cannot be as extraordinary as is implied in the simple binomial model (second version above), but it is far from small. Accounting for an n'/N on the order of 1/10 (some data cited in McLachlan (1978) imply even lower n'/N) requires a cv_p of 3. This cannot occur with a normal distribution without (forbidden) negative values. If sites have a log-normal distribution of p, one-third have a p outside of a 25-fold range about the geometric mean! A Poisson distribution among sites of numbers (n_{ca}) of local Ca^{2+} channels, with each p_o (and p, if refill is complete) proportional to the fourth power of n_{ca}, produces a cv_p of 3 only if $\langle n_{ca} \rangle$ is less than 2 and $\langle p \rangle \langle 0.02$. Counting only sites that have at least one channel, the data of Dityatev et al. (1992) for sensorimotor synapses in frog spinal cord give for cv_p a mean value of 1.35 ± 0.26, which, with the mean $\langle p \rangle = 0.200$ (±0.008, for 21 connections!) is too high to be accounted for by Poisson distribution of numbers of Ca^{2+} channels per site. It would seem therefore that release sites must differ intrinsically in their propensity to release transmitter, quite apart from possible differences in local Ca^{2+} influx per stimulus.

That refilling must be stochastic rather than fixed is a logical extension of the stochastic nature of release itself, has no implications negated by present knowledge, and is supported by observations that release remains stochastic (apparently Poisson) under conditions in which release is likely limited by "mobilization" (Elmqvist and Quastel, 1965). The resulting model is one that was given rigorous mathematical treatment by Vere-Jones (1966) and more recently considered by Melkonian (1993), whose double-barrier model is identical. It is neutral as to the physical meaning of "availability" of a quantum or how a quantum is formed and released (Melkonian, 1993). In the present context I have generalized the Vere-Jones model to one in which release sites can differ in p_o and refill rate, and these may vary in time, to obtain a method by which any hypothesis giving values to these parameters may be translated into predictions of measurable quantities in experimental data.

When refilling of sites is considered, even if it is somehow fixed rather than stochastic, incomplete refilling between stimuli is seen to imply negative covariance between outputs from successive stimuli (Vere-Jones, 1966). To my knowledge, the only published observation of such a covariance is that by Elmqvist and Quastel (1965), who found it where it is most likely to have a high value, between outputs by first and second stimuli in repeated trains. In principle, rundown in trains (paired-pulse depression) can reflect p_o's if repletion rates are low and if (perhaps hypothetical) facilitation of p_o's does not counterbalance depletion. That such rundown is obviated by low Ca^{2+}/raised Mg^{2+} (e.g., Elmqvist and Quastel, 1965; Mennerick and Zorumski, 1995) conforms with the usual views on the role of Ca^{2+} in release (i.e., in p_o), and rundown enhancement with posttetanic potentiation (Elmqvist and Quastel, 1965) suggests that p_o grows with potentiation. Apart from rundown, covariances provide in theory the only way to provide information on whether an experimentally produced change in outputs represents a change in p_o as distinct from p ($= p_o p_A$), in which p_o and p_A are inextricably mixed in means and variances. Notably, wherever n' has been reported to be much lower than the maximum m that a synapse is capable of producing (see McLachlan, 1978), significant covariance between paired-pulse responses should appear if indeed the variance/mean relationship incompatible with a low p' binomial (or Poisson) did not arise merely from error, such as neglecting the effect of time dispersion of release on the heights of synaptic signals or insufficient correction for nonlinear postsynaptic summation, which can be obviated only by a perfect (focal) voltage clamp. One important caveat about the use of covariance is that postsynaptic depression after quantal release (Mennerick and Zorumski, 1996) would lead to negative covariances between successive stimuli in much the same way as presynaptic depression arising from depletion.

From the calculations presented above covariance is unlikely to be detectable at equilibrium, with repeated stimulation, except in very large data series. However, it should be readily detectable in the autocorrelogram of miniatures if these occur at a sufficiently high frequency and most come from only a few release sites. In principle, this can provide information on refill rate(s), as well as confirm the underlying binomial model.

It is clear from the derivation of Eq. 1 that the critical assumption for the compound binomial model is not the absence of temporal or spatial variation of p_o or of p_A, all of which are in fact to be expected, but that release sites be independent. Thus, to take an extreme example, at CNS synapses an action potential may not always invade a bouton. If there is one release site per bouton, and boutons are independent with respect to noninvasion, this can have no effect on overall release statistics. However, if groups of boutons are together noninvaded or invaded (Lüscher, 1990), the release sites become nonindependent and their outputs positively correlated. In general, any positive correlation between output signals from release sites increases variance, and any negative correlation (which could be postsynaptic: mutual occlusion of signals generated on the same dendrite) reduces variances (Dityatev et al., 1992). In the case of large synapses, such as neuromuscular junctions, a positive correlation between sites would be expected unless Ca^{2+} that enters with each impulse (of necessity varying because of the stochastic nature of Ca^{2+} channel opening and closing) is effectively confined to domains that include only one release site (Quastel, Guan and Saint, 1992) and is denied by experimental data showing essentially Poisson outputs in low Ca^{2+}/raised Mg^{2+} (del Castillo and Katz, 1954a,b; Bain and Quastel, 1992a).

A major objective of much current work is to sort out the pre- and postsynaptic contributions to changes in synaptic efficacy (e.g., Korn and Faber, 1991). To accomplish this, the sine qua non is that one should be able to distinguish experimentally between changes in quantal amplitude or number ($\langle m \rangle$). At neuromuscular junctions this is easy because miniatures and quantal components of evoked signals are the same; this cannot be assumed for any cell with more than a single synaptic input in a well-defined small region. Indeed, in certain reductio ad absurdum scenarios, pre- and postsynaptic change may be indistinguishable by any criterion—a p ceasing to be zero at a site will be manifested in a manner identical to that of its postsynaptic response ceasing to be zero. However, in all but rather artificial constructs, the ratio $\langle S \rangle^2/\mathrm{var}(S)$ always changes in the same direction as m, and it is in the context of providing a simple method for determining var(S) without noise bias that the area product has been introduced. This function also includes other information extracted from the raw data, in the time partitioning of var(S), that may sometimes be ambiguous but sometimes useful. In particular, it could signal whether one is recording from a mix of synapses with different quantal responses (Redman, 1990; Walmsley, 1993; Jack et al., 1994). If not, and with important caveats, it may yield an estimate of the amplitude of the channel components of the postsynaptic response, which is also obtainable in principle from comparison of the coefficients of variation of response amplitude and area. It is also one of a group of functions, including its third-moment analog, point-to-point variance, point-to-point third moment, and, indeed, the mean of the signals, that can each be calculated easily from a data set and compared with theoretical functions, to discriminate between different models as to the make-up of the synaptic signals (e.g., Redman, 1990; Jack et al., 1994).

Information on the amplitude, time course, and variability of unit quantal synaptic events, as well as their number (m), is of course embodied in evoked synaptic signals and might be extractable with sufficiently sophisticated programs and reasonable assumptions about the distribution of p between release sites (rev. Redman, 1990; e.g., Dityatev et al., 1992). More directly, the essential problem would disappear if quanta in evoked responses could be counted unambiguously, or if one could be sure that miniatures were the same as the quantal components of evoked responses. Especially pertinent to this is the observation by Abdul-Ghani and Pennefather (1993) that synapses in cultured hippocampal neurons in the presence of Sr^{2+} produce poststimulus "after-discharges" of miniatures. It was previously shown that at the mouse neuromuscular junction this phenomenon can be attributed to presynaptic persistence of Sr^{2+} that has entered through impulse-opened Ca^{2+} channels, acting on the release system(s) in the same way as Ca^{2+} (Bain and Quastel, 1992a), except for the multiplicative component of facilitation (Bain and Quastel, 1992b). Assuming the same for hippocampal synapses, "tail" miniatures must come from boutons that have been stimulated. Moreover, they can be regarded as time-dispersed synaptic signals whose m's can be found by counting rather than complex signal analysis. A complication is that varied p values among release sites implies that the quantal mix from different sites will not be the same as in the "phasic" responses—sites with high p_o's have fewer quanta available for subsequent release by residual Sr^{2+} (Fig. 4), but if the amplitude distribution of such miniatures is invariant with time after each stimulus (or with stimulation frequency), one could be confident that they fairly represent the quantal components of the "phasic" signals. The same, of course, applies to miniatures accelerated by repetitive stimulation without Sr^{2+}, the advantage of Sr^{2+} being the plethora of miniatures obtainable at low or moderate stimulation frequencies, permitting, if release sites are associated with quantal responses uncorrelated with their p's, a direct measure of the presynaptic contribution to any change in synaptic efficacy.

APPENDIX

1. Distribution function of outputs (Table 1)

At each site this function takes the form prob(failure) = $q = 1 - p$, prob(success) = p. With one site the probability of a failure, $P_1(0) = q_1$, and the probability of obtaining a single quantum, $P_1(1) = p_1$; with two sites $P_2(0) = q_2P_1(0)$, $P_2(1) = q_2P_1(1) + p_2P_1(0)$, and the probability of having two quanta in a response, $P_2(2) = p_2P_1(1)$; with three sites $P_3(0) = q_3P_2(0)$, $P_3(1) = q_3P_2(1) + p_3P_2(0)$, $P_3(2) = q_3P_2(2) + p_3P_2(1)$, $P_3(3) = p_3P_2(2)$; and so on. Thus, for any number of sites, the output distribution is found by successive convolutions.

With respect to the number of expected failures the compound binomial gives

$$P(0) = q_1q_2q_3q_4 \cdots q_N$$

$$\ln P(0) = \ln(1 - p_1) + \ln(1 - p_2) + \ln(1 - p_3) + \cdots$$

In general,

$$\ln(1 - x) = -(x + x^2/2 + x^3/3 + \cdots)$$

With $x < 0.3$ the first two terms in this expansion give a good approximation, hence,

$$-\ln P(0) \cong p_1 + p_2 + p_3 + \cdots (p_1^2 + p_2^2 + p_3^2 + \cdots)/2$$

$$= N\langle p\rangle + N\langle p\rangle^2(1 + cv_p^2)/2$$

$$= N\langle p\rangle(1 + \langle p\rangle(1 + cv_p^2)/2)$$

The simple binomial gives, for the probability of a failure,

$$P(0) = (1 - p')^{n'}$$

$$\ln P(0) = n' \ln(1 - p')$$

$$= -[N/(1 + cv_p^2)](p' + p'^2/2 + \cdots)$$

$$-\ln P(0) \cong [N/(1 + cv_p^2)](\langle p\rangle(1 + cv_p^2) + \langle p\rangle^2(1 + cv_p^2)^2/2)$$

$$= N\langle p\rangle(1 + \langle p\rangle(1 + cv_p^2)/2)$$

Thus, provided no single p is more than about 0.3, the simple and compound binomials give the same number of failures.

For the number of unit responses, $P(1)$, the simple binomial gives

$$P(1) = n'p'(1 - p')^{n'-1}$$

$$= P(0)\langle m\rangle/(1 - p')$$

$$\cong P(0)\langle m\rangle(1 + p')$$

The compound binomial gives $P(1)$ as the sum of probabilities that any site releases one quantum while others release none:

$$P(1) = \sum_i (P(0)p_i/(1 - p_i) \cong P(0)\sum (p_i + p_i^2)$$

$$= P(0)N(\langle p\rangle + \langle p\rangle^2 + \text{var}(p)) = P(0)\langle m\rangle(1 + p')$$

That is, the expected number of unit successes is also the same for both models, to a first approximation valid for small p'.

2. Release asynchrony (Table 2)

The calculations for Table 2 used a variant on Eq. 2, which was considered as pertaining to a series of k (small) p_i's, each in a small time ($\delta t = T/k$), the output for one stimulus being the sum of the k outputs. Using these equations as they stand would require storing $k^2/2$ covariances to obtain the variance of the sum; the following approach, using equations for covariances in terms of outputs, is more economical. Writing m_i, etc., for $E(m_i)$, etc., and putting $n = 1$, for one release site, from Eq. 2:

$$\text{cov}(m_i, m_{i+1}) = m_i(p_{i+1}\alpha_i - m_{i+1})$$

$$\text{cov}(m_i, m_{i+2}) = m_i(p_{i+2}\alpha_{i+1} + p_{i+1}\alpha_i\beta_{i+1}q_{i+1} - m_{i+2}),$$

etc. These lead to a fairly simple recursive formula. Defining:

$$z_k = p_k; \qquad M_k = m - m_k$$

$$z_{k-1} = p_{k-1} + \beta_{k-1}z_kq_{k-1}; \qquad M_{k-1} = M_k - m_{k-1}$$

$$z_{k-2} = p_{k-2} + \beta_{k-2}z_{k-1}q_{k-2}; \qquad M_{k-2} = M_{k-1} - m_{k-2}$$

etc., and designating as m, the sum of outputs, $m_1 + m_2 + m_3 + .. + m_k$, variance is given by

$$\text{var}(m) = m(1 - m) + 2D$$

where

$$D = \alpha_{k-1}z_kM_k + \alpha_{k-2}z_{k-1}M_{k-1} + \alpha_{k-3}z_{k-2}M_{k-2} + \cdots$$

For any sequence of k p_o's, repeated indefinitely, the series of n_i's is given by starting with $n_1 = 1$. Then, $n_2 = \alpha_1 + \beta_1 q_1 n_1$, $n_3 = \alpha_2 + \beta_2 q_2 n_2, \ldots, n_k = \alpha_{k-1} + \beta_{k-1} q_{k-1} n_{k-1}$, $n_1 = \alpha_k + \beta_k q_k n_k$ (α_k being the last before the next stimulus), and the sequence starts again with the new value for n_1. A few iterations give equilibrium n_1, n_2, etc., each m_i is $p_i n_i$, and from the above z's and M's, variance is obtained. The covariance between outputs from sequential stimuli is obtained by using the same logic for a group of two adjacent stimuli, seen as a sequence in $2k$ time bins; the mean is $2m$ and the variance of the pair is $2\mathrm{var}(m)$ plus twice the covariance.

For Table 2 it was assumed 1) that R_A is constant and R_L and R_o are negligible, so that $\tau = 1/R_A$, giving $\alpha = 1 - \exp(-\delta t/\tau)$, for each δt, and 2) that instantaneous release rates follow a time course proportional to $(\exp(-at) - \exp(-bt))$, with $b/a = 4$. The time between stimuli, T, was divided into $k = 20{,}000$ bins of duration δt. The resulting series of release rates, $r_1, r_2, \ldots, r_i, \ldots$, results in a net $p_o = 1 - \exp(-\Sigma\, r_i \delta t)$ that is the same as would be caused by a steady "square pulse" with amplitude r_o equal to maximum r_i maintained for a time period t_{eff}. The calculations showed that $\langle m \rangle$ and $\mathrm{var}(m)$ become greater than for instantaneous release to an extent that depends upon p_o, t_{eff}, τ, and T. The summarization in Table 2 is based on the observation that changes in apparent n (n_a, defined by $\mathrm{var}(m) \equiv m(1 - m/n_a)$ and $\Delta n_a = n_a - 1$), were nearly exactly proportional to t_{eff}. Changes in apparent p (defined by $p_a \equiv 1 - \mathrm{var}(m)/m$ and $\Delta p_a = p_a - p$) were negative; $\Delta p_a/\Delta n_a$ was nearly linearly related to p_o. The tabulated numbers give changes in n_a and p_a (from 1.0 and p, respectively) that together give the change in variance and m (= mean output = $n_a p_a$). For example, if the refill rate is 2/s, τ is 500 ms; stimulation at 0.5 Hz gives $T/\tau = 4.0$; if t_{eff} is 10 ms, t_{eff}/τ is 2%. For a site with $p_o = 0.99$ (last item in Table 2), n_a (apparent n) is increased from 1.0 by 2.5 × 2% (= 5%) to 1.05; p_a is decreased by the formula by (0.99–0.54) × 5% (= 2.23%), from 0.972 to 0.950. With instantaneous release m would be 0.972 and the variance would be $0.972 \times (1 - 0.972) = 0.027$; with release spread in time m becomes 1.0 and variance becomes $1.05 \times 0.95 \times (1 - 0.95) = 0.050$. For the same parameters a site with $p_o = 0.5$ would have n_a increased from 1.0 to 1.018; and p_a decreased from 0.495 by $(0.99 - 0.55 \times 0.5) \times 1.8\%$, to 0.489.

Varying the ratio b/a altered the results to some extent; with $b/a = 1.1$ changes in n_a were 16% lower than tabulated, and with $b/a = 10$, changes in n_a were 18% higher than tabulated; equations relating Δp_a to Δn_a were virtually the same.

The changes in covariances produced by release asynchrony (not listed) were rather complex but relatively small, except when covariances were so small as to be below any possibility of experimental detection. However, if a high release probability persists for an appreciable time between stimuli, there are substantial negative covariances between outputs within this time.

3. Coefficients of variation of heights and areas

Suppose we have a series of k records, in each of which a certain number of channels are open (sometimes 0, sometimes 1, sometimes 2, etc.), with relative frequencies f_0, f_1, f_2, etc., adding up to unity, so the number with n channels that open is kf_n. Moments for record area (S) can be calculated in much the same way as for Eq. 4. Each record with n open channels is expected to contribute on average nS_1 to the overall sum of areas and $(nS_1)^2 + nV_1$ to the overall sum of squares, where S_1 and V_1 denote the average area and within-group variance of records with just one channel. Thus,

$$E(S) = \mu_1' = \sum (kf_n nS_1)/k = S_1 \sum f_n n$$

$$\sum (S^2)/k = \mu_2' = \sum (kf_n(nS_1)^2)/k + \sum (kf_n nV_1)/k$$

$$= S_1^2 \sum (f_n n^2) + V_1 \sum (f_n n)$$

$$\mathrm{var}(S) = \mu_2' - (\mu_1')^2 = S_1^2 \sum f_n n^2 + V_1 \sum f_n n - S_1^2 (\sum f_n n)^2$$

$$cv_s^2 = \mu_2'/(\mu_1')^2 - 1$$

$$= \sum f_n n^2/(\sum f_n n)^2 + (V_1/S_1^2)/\sum f_n n - 1,$$

where the summation is over all n. With exponentially distributed channel durations $S_1 = h_c\tau_c$ and $V_1 = (h_c\tau_c)^2$. Moreover, $\Sigma f_n n = \langle n \rangle$, the mean number of open channels per response, and $\Sigma f_n n^2/(\Sigma f_n n)^2 - 1$ is cv_n^2. Also, if channels open simultaneously, $\mathrm{var(height)}/\langle \mathrm{height} \rangle^2 = cv_H^2 = cv_n^2$. Therefore,

$$cv_s^2 = cv_n^2 + 1/\langle n \rangle (= cv_H^2 + 1/\langle n \rangle)$$

however n is distributed. Using similar logic,

$$S3/\langle S \rangle^3 = N3/\langle n \rangle^3 + (3/\langle n \rangle)cv_n^2 + 2/\langle n \rangle^2$$

where $N3$ (note $H3/\langle H \rangle^3 = N3/\langle n \rangle^3$) and $S3$ are third moments about means. To obtain this one uses $S3 = 2(h_c\tau_c)^3$ for records with just one channel.

It is notable that for Poisson distributed n, as might be expected in a multichannel patch where a small fraction of channels open and then close in response to repeated constant stimuli, $E(n) = \mathrm{var}(n) = N3$, and therefore $\langle S \rangle = \langle n \rangle h_c\tau_c$, $\mathrm{var}(S) = 2\langle n \rangle (h_c\tau_c)^2$ and $S3 = 6\langle n \rangle (h_c\tau_c)^3$.

4. The area product

i. Numerical equivalence of area product sum and var(S)

We have k records, each with K values. To avoid double subscripts, let us say that the first record consists of values $a_1, a_2, \ldots, a_K$ with sum S_a, the next $b_1, b_2, \ldots, b_K$ with sum S_b, etc. By definition,

$$k\langle y_j \rangle = a_j + b_j + c_j + \cdots$$

and

$$k\langle S \rangle = S_a + S_b + S_c + \cdots = k \sum \langle y_j \rangle,$$

where the summation is over all K values of j.

Each A_j multiplied by k is

$$kA_j = (S_a - \langle S \rangle)(a_j - \langle y_j \rangle) + (S_b - \langle S \rangle)(b_j - \langle y_j \rangle) + \cdots$$

$$= (S_a a_j + S_b b_j + S_c c_j + \cdots) - \langle S \rangle (a_j + b_j + c_j + \cdots)$$

$$- \langle y_j \rangle (S_a + S_b + S_c + \cdots) + k\langle y_j \rangle \langle S \rangle$$

$$= (S_a a_j + S_b b_j + S_c c_j + \cdots) - \langle S \rangle k \langle y_j \rangle - \langle y_j \rangle k \langle S \rangle$$

$$+ k\langle y_j \rangle \langle S \rangle$$

$$= (S_a a_j + S_b b_j + S_c c_j + \cdots) - k\langle S \rangle \langle y_j \rangle.$$

Summing over the K A_j's,

$$k \sum A_j = S_a(a_1 + a_2 + \cdots + a_K) + S_b(b_1 + b_2 + \cdots . + b_K)$$

$$+ S_c(c_1 + c_2 + \cdots + c_K) \cdots - k\langle S \rangle^2$$

$$= \sum S^2 - k\langle S \rangle^2 = k\,\mathrm{var}(S).$$

Because the above equations involve actual numbers, not expected values, $\Sigma A_j = \mathrm{var}(S)$ is therefore a numerical identity, whatever the values in the k groups of K numbers.

ii. Linear growth of A_j with j with slope h_c

Assume the same model and terminology as in (3) above, with channels opening simultaneously. To calculate A_j's we first obtain ΣSy_j for records with any distribution of numbers of 0, 1, 2, etc. channels opening in each. Because in general, $A_j = \langle Sy_j \rangle - \langle S \rangle \langle y_j \rangle$, the expected contribution of a group of records to overall $\Sigma\, Sy_j$ is the number of records in the group multiplied by its $\langle Sy_j \rangle + A_j$. Analogous to variance, if we know A_j for records with just one channel, say A_{1j}, we can simply write for the group of records, each with n open channels, that its A_j is nA_{1j}; this follows from (i) above, because the average time course of signals is the same for any n. Furthermore, for records with one channel, $\langle S\rangle = h_c\tau_c$ and $\langle y_j\rangle = h_c\exp(-j\delta t/\tau_c)$. For brevity designate these S_1 and y_{1j}, respectively. Then, for records with just n channels, $\langle S\rangle = nS_1$ and $\langle y_j\rangle = ny_{1j}$

For all of the k records, summing over all n,

$$\sum Sy_j/k = A_{1j}\sum f_n n + S_1 y_{1j}\sum f_n n^2.$$

Now $\Sigma f_n n$ is $\langle n\rangle$, the mean number of channels opening per response, and $\Sigma f_n n^2$ is $\mathrm{var}(n) + \langle n\rangle^2$. Thus, $\langle S\rangle = \langle n\rangle S_1$, $\langle y_j\rangle = \langle n\rangle y_{1j}$, and, equating means and expected values,

$$\begin{aligned} A_j &= \langle Sy_j\rangle - \langle S\rangle\langle y_j\rangle \\ &= A_{1j}\langle n\rangle + S_1 y_{1j}(\mathrm{var}(n) + \langle n\rangle^2) - (nS_1)(ny_{1j}) \\ &= A_{1j}\langle n\rangle + S_1 y_{1j}\,\mathrm{var}(n) \end{aligned}$$

$$A_j/\langle y_j\rangle = A_{1j}/y_{1j} + S_1\,\mathrm{var}(n)/\langle n\rangle. \qquad \text{(A1)}$$

We must now find A_{1j}, or rather, $A_1(t)$. Consider records at which one channel opens at 0 time, and a sample time δt that can be as small as we wish. The probability that a channel is still open at time δt is $1 - \exp(-\delta t/\tau_c)$, the probability it is still open at time $2\delta t$ is $1 - \exp(-2\delta t/\tau_c)$, etc. Let $a = \exp(-\delta t/\tau_c)$. Then the probability that the channel is open for just one point (or less) is $1 - a$, for just 2 points $(1 - a^2) - (1 - a) = a(1 - a)$, . . . for x points $a^{x-1}(1 - a)$, and these are associated with sums (S) h_c, $2h_c$, . . . , xh_c, respectively, and with respective cross-products (y_iS) h_c^2 only at point 1, $2h_c^2$ at points 1 and 2 only, . . . , xh_c^2 for all points $\leq x$ and elsewhere 0. Hence, summing to infinity products of value and the probability of that value occurring,

$$\begin{aligned} \langle y_1S\rangle/h_c^2 &= (1 - a)(1 + 2a + 3a^2 + 4a^3 + \cdots) \\ &= 1/(1 - a) \end{aligned}$$

$$\begin{aligned} \langle y_2S\rangle/h_c^2 &= (1 - a)(2a + 3a^2 + 4a^3 + \cdots) \\ &= a/(1 - a) + a \end{aligned}$$

$$\langle y_3S\rangle/h_c^2 = (1 - a)(3a^2 + 4a^3 + \cdots) = a^2/(1 - a) + 2a^2$$

$$\langle y_iS\rangle/h_c^2 = a^{i-1}/(1 - a) + (i - 1)a^{i-1}.$$

By the same logic,

$$\langle y_i\rangle/h_c = (1 - a)(a^{i-1} + a^i + a^{i+1} + \cdots) = a^{i-1}$$

and

$$\begin{aligned} \langle S\rangle/h_c &= \langle y_1\rangle + \langle y_2\rangle + \langle y_3\rangle + \cdots = 1 + a + a^2 + a^3 + \cdots \\ &= 1/(1 - a), \end{aligned}$$

hence

$$\begin{aligned} A_i &= \langle y_iS\rangle - \langle y_i\rangle\langle S\rangle \\ &= h_c^2 a^{i-1}[1/(1 - a) + i - 1 - 1/(1 - a)] \\ &= h_c^2 a^{i-1}(i - 1) \end{aligned}$$

and

$$A_i/\langle y_i\rangle = h_c(i - 1).$$

Taking limits as $\delta t \rightarrow 0$, and taking into account that all time summations are multiplied by δt, we obtain

$$S_1 = h_c\tau_c; \qquad A_1(t)/y_1(t) = h_c t$$

and, substituting into Eq. A1,

$$A(t)/\langle y(t)\rangle = h_c t + h_c\tau_c\,\mathrm{var}(n)/\langle n\rangle$$

or

$$A_j/\langle y_j\rangle = jh_c + h_c\tau_c\,\mathrm{var}(n)/\langle n\rangle,$$

the latter pertaining if summations (time integrals) use the digitizing interval as the time unit and τ_c is expressed in these units. Thus, $A_j/\langle y_j\rangle$ rises linearly with j, with slope h_c, for any distribution of numbers of channels per response. For the situation in Fig. 6 *B*, where a patch with a single channel is envisaged, and p is the probability of it opening, $\langle n\rangle = p$ and $\mathrm{var}(n) = p(1 - p)$.

5. Practical considerations in determining channel height from the area product

To determine how well channel height (h_c) might be determined from the area product, I have made long series of simulations (at least 50 in each group) with various combinations of channel time constant (τ_c), channels per quantum (n_c), mean quantal content ($\langle m\rangle$), number of records analyzed together (k), time constants governing dispersion of channel opening, and noise at various amplitudes. These were all found to interact in a rather complex way to influence the accuracy and variability of putative h_c. The results are summarized below.

Finding h_c

If channels all open simultaneously, theory (see (4) above) gives $A_j/\langle y_j\rangle = a + bj$, where a is what $A_j/\langle y_j\rangle$ would be if all quanta had a uniform time course and b corresponds to channel height, h_c. In principle, a and b can be found in two ways: 1) linear correlation of $A_j/\langle y_j\rangle$ with j, or 2) fitting of A_j to the equation $A_j = a\langle y_j\rangle + b(j\langle y_j\rangle)$, in each case by least squares, and after subtracting average prestimulus values from A_j and $\langle y_j\rangle$. In practice, it was found that with method 1) values of b best corresponded to true h_c when all points were weighted by $\langle y_j\rangle^3$, and this gave results identical to those of method 2) with weighting by $\langle y_j\rangle$. Points used were from $\langle y_j\rangle$ at 90% of its peak to 2.5% of its peak, to exclude (most) of the error otherwise arising from (minor) time dispersion of channel opening. Invariably, this method gave less bias and less variability of putative h_c than simply using coefficients of variation of signal heights and areas (see (3) above).

Noise and quantal visibility

All simulated records were constructed to correspond to signals plus Gaussian noise as they would appear using a low-pass filter at a frequency (f) corresponding to one-half the *A–D* conversion rate, to avoid aliasing. It is convenient to define Z as the baseline noise standard deviation ("rms noise"), when records are recorded/filtered in this way, divided by channel height, h_c. Then a simple rule of thumb is that a (nearly) exponentially decaying signal such as a quantal response becomes indistinguishable from random noise fluctuations by eye or by cross-correlation with a template, when Z is about $n_c(\tau_c/80)^{0.5}$, where n_c is the mean number of channels per quantum and τ_c is expressed in time units equal to the digitizing interval. This applies whatever apparent improvement can be produced by filtering with lower f. Thus, in the simulation in Fig. 7 where "unfiltered" rms noise was 0.5 units, unitary G1 quanta (mean size 1 unit, τ_c = 40 ms or points) are visible but G2 unit quanta (mean peak amplitude = 0.167 units, τ_c = 120 points) cannot be seen in the records with added noise in Fig. 7 *A*, despite (for the illustration) a low-pass 25-Hz filter.

Filtering the mean and area product

Although smoothing individual records was not useful, smoothing of both $A(t)$ and $\langle y(t) \rangle$ by simple low-pass filtering before finding b by least squares led to less variability of b. The time constant for this was chosen as 10% of the apparent time constant by which $\langle y_j \rangle$ decayed from 90% to 10% of its peak; this had a negligible effect to reduce b (see below). Any further smoothing with this or smaller time constants, or using running averages, had a negligible effect.

Uncertainty in h_c

In the absence of added noise the standard deviation of b/h_c (which was apparently normally distributed) was always within about 20% of the value given by the following empirical formula:

$$\mathrm{SD}_o(b/h_c) = (1 + 0.75/\tau_c)[1 + 1.2(n_c w)^{0.5}]/k^{0.5},$$

where $w = (1 - p' + cv_q^2)$, cv_q being the coefficient of variation of quantal height. Alternatively, in terms of a, $n_c w = a/(h_c \tau_c)$. Thus $\mathrm{SD}_o(b/h_c)$ can be quite appreciable if n_c is not fairly small and k is only a few hundred. For example, in the simulation in Fig. 7 C ($k = 400$), with quanta having $\tau_c = 40$ and $p' = 0.33$, the variation of b/h_c is about ± 0.2 with $n_c = 10$, and ± 0.4 with $n_c = 50$.

In the presence of noise, the standard deviations of b/h_c fit well to the formula

$$\mathrm{SD}(b/h_c) = (V_o + V_n)^{0.5},$$

where V_o is $[\mathrm{SD}_o(b/h_c)]^2$, and V_n, within about 30%, is given by

$$V_n = (15/k)Z^{(2+x)} w^{0.5} \langle m \rangle^{-1.25} \tau_c^{-0.75}.$$

Here $x = 1.25 \exp(-\tau_c n_c/125)$. For example, in the simulation for Fig. 7 C the noise making G2 quanta invisible raises the $\mathrm{SD}(b/h_c)$ for these quanta from 0.13 to 0.16 when n_c is 2 and from 0.2 to 0.4 when $n_c = 10$. The tendency for V_n to increase with reduced τ_c results in no gain in accuracy of determining h_c by reducing the sampling rate and, with it, the prefilter frequency (f); in the other direction no gain in accuracy occurred with increasing sampling rate, to make τ_c more than about 40 points.

Systematic error produced by noise

Without noise, b was an unbiased estimator of h_c, except (see below) with time dispersion of channel opening or filtering. However, noise produced a downward bias that was evident when $\langle m \rangle$ and/or k was low. Empirically:

$$b/h_c = \exp[-(450/k)Z^2/(n_c \tau_c \langle m \rangle)].$$

Thus, for example, with quanta composed of an average of 10 channels, a noise level that just makes individual quanta invisible ($Z = 10$ if $\tau_c = 80$) gives with $\langle m \rangle = 1$, $b/h_c = 0.57$ at $k = 100$ and $b/h_c = 0.87$ with $k = 400$; with $\langle m \rangle = 2.5$, $b/h_c = 0.80$ and 0.96 at $k = 100$ and 400, respectively. Such error is signaled by the b obtained from using all records together being higher than the average of b's obtained by analyzing subgroups of records separately.

Systematic error due to filtering or time dispersion

If channels do not all open simultaneously, $\langle Sy(t) \rangle$ and $\langle y(t) \rangle$ are convoluted by the time course of channel opening; exponentially distributed lags of channel opening with a mean of τ_d have the same effect as a low-pass filter with time constant τ_d applied to the records. For any particular "filter" (including, for example, electrotonic conduction) the net effect can be calculated by setting up theoretical A_j's and $\langle y_j \rangle$'s, calculating $\langle Sy_j \rangle$'s, convoluting the two latter by the appropriate function, and recalculating the resulting A_j's and $A_j/\langle y_j \rangle$'s. Using this method or simulation with time-dispersed channel opening (with up to three time constants) gave the same results:

1. Effects to reduce b are less than 2% if the largest time constant, τ_d, is less than about 20% of τ_c. Thus, in Fig. 7 C, where quanta of both G1 ($\tau_c = 40$) and G2 ($\tau_c = 120$) type were assumed released with lag time constants of 2 and 4.5 points, and channel opening had mean lags of 4 ms and 20 ms respectively, these parameters could be expected to have little effect on b.

2. With higher τ_d the graph of $A_j/\langle y_j \rangle$ versus j becomes curvilinear, curving upward to a slope of h_c when $\tau_d < \tau_c$ and curving down to a slope of 0 when $\tau_d > \tau_c$. With b determined as described above (fitting to a linear relation that is no longer true), b/h_c averages 0.86 at $\tau_d/\tau_c = 0.5$, 0.68 at $\tau_d/\tau_c = 0.75$, and 0.5 at $\tau_d = \tau_c$; with higher τ_d, b/h_c declines approximately e-fold for each increment of τ_d by τ_c.

This work was supported by a grant from the Muscular Dystrophy Association of Canada.

REFERENCES

Abdul-Ghani, M. A., and P. S. Pennefather. 1993. Sr^{2+} afterdischarge and quantal analysis of synaptic transmission between pairs of cultured mouse hippocampal neurons. *Biophys. J.* 64:A236.

Bain, A. I., and D. M. J. Quastel. 1992a. Quantal transmitter release mediated by strontium at the mouse motor nerve terminal. *J. Physiol. (Lond.).* 450:63–87.

Bain, A. I., and D. M. J. Quastel. 1992b. Multiplicative and additive Ca^{2+}-dependent components of facilitation at mouse endplates. *J. Physiol. (Lond.).* 455:383–405.

Brown, T. H., D. H. Perkel, and M. W. Feldman. 1976. Evoked neurotransmitter release: statistical effects of nonuniformity and nonstationarity. *Proc. Natl. Acad. Sci. USA.* 73:2913–2917.

del Castillo, J., and B. Katz. 1954a. Quantal components of the end-plate potential. *J. Physiol. (Lond.).* 124:560–573.

del Castillo, J., and B. Katz. 1954b. Statistical factors involved in neuromuscular facilitation and depression. *J. Physiol. (Lond.).* 124:574–585.

del Castillo, J., and B. Katz. 1956. Biophysical aspects of neuro-muscular transmission. *Prog. Biophys.* 6:121–170.

Dityatev, A. E., V. M. Kozhanov, and S. O. Gapanovich. 1992. Modeling of the quantal release at interneuronal synapses: analysis of permissible values of model moments. *J. Neurosci. Methods.* 43:201–214.

Elmqvist, D., and D. M. J. Quastel. 1965. A quantitative study of end-plate potentials in isolated human muscle. *J. Physiol. (Lond.).* 178:505–529.

Hubbard, J. I. 1963. Repetitive stimulation at the mammalian neuromuscular junction and the mobilisation of transmitter. *J. Physiol. (Lond.).* 169:641–662.

Jack, J. J. B., A. U. Larkman, G. Major, and K. J. Stratford. 1994. Quantal analysis of the synaptic excitation of CA1 hippocampal pyramidal cells. *In* Molecular and Cellular Mechanisms of Neurotransmitter Release. L. Stjärne, P. Greengard, S. E. Grillner, T. G. M. Hökfelt, and D. R. Ottoson, editors. Raven Press, New York.

Katz, B., and R. Miledi. 1965. The measurement of synaptic delay and the time course of acetylcholine release at the neuromuscular junction. *Proc. R. Soc. Lond. B.* 161:483–495.

Korn, H., and D. S. Faber. 1991. Quantal analysis and synaptic efficacy in the CNS. *Trends Neurosci.* 14:439–445.

Lüscher, H. R. 1990. Transmission failure and its relief in the spinal monosynaptic reflex arc. *In* The Segmental Motor System. M. D. Binder and L. M. Mendell, editors. Oxford University Press, New York. 328–348.

Liley, A. W., and K. A. K. North. 1953. An electrical investigation of effects of repetitive stimulation on mammalian neuromuscular junctions. *J. Neurophysiol.* 16:509–527.

Martin, A. R. 1955. A further study of the statistical composition of the end-plate potential. *J. Physiol. (Lond.).* 130:114–122.

McLachlan, E. M. 1978. The statistics of transmitter release at chemical synapses. *In* International Review of Physiology, Neurophysiology III, Vol. 17. R. Porter, editor. University Park Press, Baltimore. 49–117.

Melkonian, D. S. 1993. Transient analysis of a chemical synaptic transmission. *Biol. Cybern.* 68:341–350.

Mennerick, S., and C. F. Zorumski. 1995. Paired-pulse modulation of fast excitatory synaptic currents in microcultures of rat hippocampal neurons. *J. Physiol. (Lond.).* 488:85–101.

Mennerick, S., and C. F. Zorumski. 1996. Postsynaptic modulation of NMDA synaptic currents in rat hippocampal microcultures by paired-pulse stimulation. *J. Physiol. (Lond.).* 490:405–417.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1992. Numerical Recipes in C, 2nd ed. Cambridge University Press.

Quastel, D. M. J., Y.-Y. Guan, and D. A. Saint. 1992. The relation between transmitter release and Ca^{2+} entry at the mouse neuromuscular junction: role of stochastic factors causing heterogeneity. *Neuroscience.* 51: 657–671.

Redman, S. 1990. Quantal analysis of synaptic potentials in neurons of central nervous system. *Physiol. Rev.* 70:165–198.

Vere-Jones, D. 1966. Simple stochastic models for the release of quanta of transmitter from a nerve terminal. *Aust. J. Statist.* 8:53–63

Walmsley, B. 1993. Quantal analysis of synaptic transmission. *In* Electrophysiology: A Practical Approach. D. I. Wallis, editor. IRL Press, Oxford.

Weatherburn, C. E. 1961. A First Course in Mathematical Statistics. Cambridge University Press, Cambridge.