J. theor. Biol. (2000) **207,** 195–216
doi:10.1006/jtbi.2000.2164, available online at http://www.idealibrary.com on IDEAL®

Designing Lymphocyte Functional Structure for Optimal Signal Detection: *Voilà*, T cells

ANDRÉ J. NOEST*

Department of Theoretical Biology, Utrecht University, Padualaan 8, *NL*-3584-*CH Utrecht, The Netherlands*

(*Received on* 14 *October* 1999, *Accepted in revised form on* 8 *August* 2000)

One basic task of immune systems is to detect signals from unknown "intruders" amidst a noisy background of harmless signals. To clarify the functional importance of many observed lymphocyte properties, I ask: What properties would a cell have if one *designed* it according to the theory of optimal detection, with minimal regard for biological constraints? Sparse and reasonable assumptions about the statistics of available signals prove sufficient for deriving many features of the optimal functional structure, in an incremental and modular design. The use of one common formalism guarantees that all parts of the design collaborate to solve the detection task. Detection performance is computed at several stages of the design. Comparison between design variants reveals e.g. the importance of controlling the signal integration time. This predicts that an appropriate control mechanism should exist. Comparing the design to reality, I find a striking similarity with many features of T cells. For example, the formalism dictates clonal specificity, serial receptor triggering, (grades of) anergy, negative and positive selection, co-stimulation, high-zone tolerance, and clonal production of cytokines. Serious mismatches should be found if T cells were hindered by mechanistic constraints or vestiges of their (co-)evolutionary history, but I have not found clear examples. By contrast, fundamental mismatches abound when comparing the design to immune systems of e.g. invertebrates. The wide-ranging differences seem to hinge on the (in)ability to generate a large diversity of receptors.

1. Introduction: Designing Lymphocytes from Scratch

The structural and functional complexity of vertebrate immune systems [see Janeway & Travers (1996) for a broad overview] should reflect the need for an effective defence against a large and changing set of pathogens. Experimental immunology is revealing a vastly detailed view of the machinery behind many functionally vital features, but the size and complexity of this information creates substantial obstacles to reaching a coherent and quantitative understanding of how the many parts of the machinery jointly produce the overall functional behaviour. Data-driven, bottom-up modelling allows a quantitative study of specific parts of the system, but it is not geared to addressing integrative aspects. Top-down models have often been used in immunology to fill this gap, and they have provided the field with some of its central notions (Langman & Cohn, 1996): prime examples are Burnet's clonal selection theory, and the

*E-mail: a.j.noest@bio.uu.nl

0022–5193/00/220195 + 22 $35.00/0

Bretscher-Cohn "invention" of T-helper cells. However, the verbal nature of most of these models limits one's ability to check their consistency, evaluate their functional behaviour, and compare them quantitatively to alternative models or to reality.

In this paper, I follow a slightly unusual approach which is both top-down and quantitative: using the formalism of statistical detection theory, I derive specifications for a large set of functionally coherent features which a cell should possess if it is to perform its task, which I assume to be near-optimal detection of unknown intruders amidst a noisy background. Note how this "design" approach differs from the more usual models, in which one starts by assuming some functional structure (with free parameters), and then derives its behaviour, perhaps using the result to optimize the parameters. Here, I start from one (deceptively) simple formula which describes optimal detection very generally, add some assumptions on the statistics of the available signals, and derive the optimal functional structure. The resulting design, having a few free parameters, can then be analysed for its performance in a more usual way.

In deriving the design, I avoid relying on specifically biological or immunological ideas or experimental data. Thus, the resulting functional features can be compared to those of real immune systems without an inherent bias: if the design matches with reality, this cannot be a trivial consequence of logical circularity, as is a distinct risk in data-driven modelling. Indeed, the design can be used as a structured, quantitative "reference" or "null-model" for any cell-type with which it shares its task and signal structure. One may guess that T cells (especially "helper"-T cells) fit such a role, but this is no more than an invitation to compare the design properties to those of real T cells. Conversely, finding mismatches between design and reality (at the functional level) could be useful as indicators of where real T cells suffer from mechanistic or historical constraints. An optimal design can also be used to determine the limit to detection performance for real lymphocytes, or alternative models of them, which share the same task and signal structure. Indeed, I will compute the performance of some moderately equipped versions of the design.

From a somewhat broader viewpoint, the design allows one to address questions which would otherwise remain within the realm of indeterminate speculation—for example, Which features are essential for approaching ideal performance? How does the functional structure depend on the properties of the available signals? What if a variant scheme were used? Which, if any, real features might reflect the contingencies of evolutionary history, rather than the selective advantage of improved detection? Do any real features indicate or contradict that (co-)evolution has honed the system to near-perfection, or conversely, that it has led immunity into a morass of intricate anti-subversion measures, away from the demands of the basic task of detection? Such questions cannot make much sense, unless one can compare real immune systems to an independent reference, designed only for the basic task, and having a fully understood structure and behaviour.

Any design starts by defining a task. Here I assume it to be near-optimal detection of unpredictable "intruder"-signals, amidst a wide variety of "background"-signals. The system does not know *a priori* what types of signals may exist, nor the source (background or intruder) of any signal. Indeed, both sources may even contribute to the same type of signal. Except for one feature of the design, the task is defined as independently detecting each of a large set of possible signals. The precise nature of the response triggered by a detection can be left undefined here.

This "immune surveillance" task bears a striking resemblance to the well-studied engineering problem of detecting (e.g. electronic) signals of unknown targets among background noise or interference (Whalen, 1971): the goal is to detect even weak intruders with high probability, while keeping "false alarms" very rare. The generality of the statistical theory underlying all (near-)optimal detection systems allows the present design to borrow some basic notions and analytical techniques from engineering. The analogy can be useful only at the functional level, given the very different physical machinery which implements the scheme.

For the sake of readability, I will often use biological terms (e.g. "T cell, ligand, receptor") to denote functional units in the design, even when

their specification has not yet been derived. These are no more than convenient labels, prompted by a lack of neutral terms which are as easily memorized. The design itself does not rely on experimental data on real T cells, etc. Indeed, the design units are initially "empty", and acquire their features step by step, as dictated by added or modified assumptions about the signal statistics.

To emphasize that the basic task of detection imposes a large set of features, I ignore the distinctions between various effector-functions (cytotoxic or antibody responses, etc.). As will emerge at many points in the paper, T-helper lymphocytes match the design strikingly well, which may reflect their task in immune "surveillance" for new intruders. Cytotoxic T and B cells face roughly a similar detection task, but other demands on them, e.g. imposed by their effector-tasks, could cloud the issue I focus on here.

I also ignore "preprocessing" of intruders (breaking them down and presenting fragments as signals for possible detection). This task is vitally important, but quite distinct in its implications. I simply assume that antigen-presenting cells (APC) allow T cells to sample a wide variety of signals (ligands). If pathogens can sabotage this presentation, it cannot be undone at the detection stage, underlining the natural separation between the two tasks.

As a preview of the design I will derive in this paper, it may be useful to summarize several design features, which turn out to match well with data on real T cells:

- Clonal diversity and receptor specificity, with only one receptor type per cell.
- "Serial triggering" of many receptors per ligand (Valitutti *et al*., 1995).
- T cell activation by comparing the result of "TCR-counting" (Rothenberg, 1996) to a high threshold, which can be adjusted by co-stimulation (Viola & Lanzavecchia, 1996).
- Agonist ligands may become antagonists after a single mutation.
- Positive and negative selection in the thymus (Margulies, 1996; Alam *et al*., 1996), plus a much reduced sensitivity ("anergy") of potentially self-reactive cells.
- Reduced response to extremely strong inputs ("high-zone tolerance").
- Antigen-specific decisions to produce responses which act non-specifically (cytokines).

The layout of this paper is as follows: first, I derive a set of functional properties which make up a basic detector unit. Next, I compute its detection performance analytically, and then extend this to more practical variants. Finally, I derive several additional features which extend the design, or make it robust to imperfections in signal presentation, or to "outliers" in the background signal.

2. Deriving a Basic Functional Design

2.1. DETECTION THEORY AS GUIDING PRINCIPLE

To introduce the basics of detection theory in their simplest form, I assume temporarily that only a single known type of intruder may occur. Let P_d be the detection probability while the intruder is indeed present, and let P_f be the probability of "false alarm", i.e. of a detection occurring while the intruder is actually absent. The time-scale over which these probabilities are defined should be roughly the typical contact-time between a T cell and a suitable antigen-presenting cell (APC)—of the order of several hours (Lanzavecchia *et al*., 1999).

"Optimizing" detection performance has to somehow combine maximizing P_d with minimizing P_f, since any occurrence of false alarms on a background signal due to self-components would risk starting an autoimmune disease. Perfect detection would be trivial if background signals were constants: a threshold set infinitesimally above the background level would keep $P_f = 0$, while the arrival of any intruder signal would be detected with $P_d = 1$. In fact, backgrounds are always "noisy", and it is this which makes detection a non-trivial task. Thus, any detection system suffers from errors of two types: besides "false-alarms", with probability P_f, there will be "missed detections", with probability $1 - P_d$.

The demand to optimize both error types does not imply that one can only derive the optimal detector after arbitrarily choosing some explicit rule for "trading-off" one error type against the other. Indeed, the classical Neyman–Pearson test

(e.g. Whalen, 1971) specifies the detection scheme which maximizes P_d at *any* given P_f. Thus, the design of the optimal detector is actually independent of how one chooses to fix the one-parameter freedom of trading off the two errors. The formal free parameter, P_f, can be reparameterized conveniently by a threshold θ. Explicitly, the Neyman–Pearson detection condition then is

$$R(\{s_i\}) \equiv \frac{p^+(\{s_i\})}{p^0(\{s_i\})} > \theta_R, \qquad (1)$$

where θ_R is a suitable threshold, $\{s_i\}$ is the set of available signals, with $p^+(\{s_i\})$ and $p^0(\{s_i\})$ as their joint probability densities, respectively, with (p^+) and without (p^0) the intruder being present. Note that the likelihood ratio $R(\{s_i\})$ determines how any available "data" $\{s_i\}$ should be processed, before comparison to a threshold θ_R. Thus, the functional design of the detection machinery is encoded by the function $R(\{s_i\})$. All that follows is a step-wise decoding of this very compact and general prescription. Its functional implications can be made fully explicit under simple assumptions about the statistics of the signals $\{s_i\}$.

Given a definite function $R(\{s_i\})$, its values will have some probability density, say $p_R^{+/0}(R)$, with the usual (+/0) labelling of cases with/without intruder. The required threshold θ_R can then be determined from the definition $P_f = \int_{\theta_R}^{\infty} p_R^0(R)$. Thus, θ_R depends on the choice of P_f, and on the background density $p^0(\{s_i\})$, which generates the density $p_R^0(R)$. This already implies an important general feature: the system must use some form of "adaptation" to its background, since the full $p^0(\{s_i\})$ cannot be assumed to be known to the system *a priori*.

It is useful to note from eqn (1) that detection is invariant under all monotonic transformations on R, as long as θ_R is transformed in the same way. Thus, all realizations which differ only by such monotonicity are functionally identical, no matter how different they may appear experimentally. This vast but precisely delimited freedom of implementation may be exploited by evolution, as it is by human designers.

As a reasonable first approximation, we may assume statistical independence of s_i, since distinct chemicals tend to have distinct sources of fluctuation (In Section 4.2, I derive several consequences of a more realistic assumption in which "similar" signals are correlated). Independence means factorization of densities, $p(\{s_i\}) = \prod_i p_i(s_i)$, which suggests transforming the test quantity to $\lambda = \log R$, with threshold $\theta = \log \theta_R$. The test is now in a form which is easier to compute (in most realizations), namely a sum of terms which depend on just one signal s_i each:

$$\lambda = \sum_i [\log p_i^+(s_i) - \log p_i^0(s_i)] > \theta. \qquad (2)$$

The basic notions of detection, as just introduced, are illustrated in Fig. 1, using a test of the general form of eqn (2). It shows how P_f and P_d, which quantify detection performance, are related to each other via the probability density of λ. This, in turn, is determined by the densities of the input signals s_i via eqn (2).

For explicit calculations, $p_i^+(s_i)$ needs to represent not just the mere presence of an intruder, but how strongly it contributes to signal s_i. Thus, I replace $p_i^+(s_i)$ by a density $p_i(\rho_i; s_i)$, with continuous parameter ρ_i, which is the "signal-to-noise" ratio (SNR) of s_i. The background then has density $p_i(0; s_i)$.

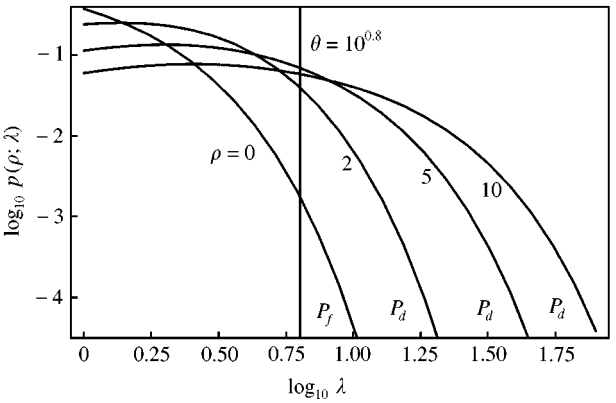

FIG. 1. A basic illustration of detection statistics: a test quantity λ, derived from noisy signals s_i, is compared to a threshold θ. Without any "intruder" (i.e. for SNR $\rho = 0$), the statistics of λ is described by its "background" density $p(0; \lambda)$. Its upper tail causes "false-alarms", at a (small) rate $P_f = \int_\theta^\infty p(0; \lambda)\,d\lambda$ set by θ. When an intruder appears, ρ increases, causing the density $p(\rho; \lambda)$ to shift to the right and change shape. This increases the detection rate $P_d = \int_\theta^\infty p(\rho; \lambda)\,d\lambda$. Useful detection performance (say $P_d > 0.5$, while P_f is many orders smaller) usually requires a considerable SNR ρ.

2.2. TRANSLATING THE FORMAL TEST INTO A FUNCTIONAL DESIGN

The formalism of Neyman–Pearson testing describes optimal signal detection. Real immune systems should kill pathogens, but it seems not too unrealistic to single out the detection of "unusual" ligand signals as a reasonably distinct subtask. This assumption motivates me to base the design on the Neyman–Pearson test formalism. Indeed, detection should be a critical step in killing pathogens, unless the effector stages of an immune response are very unreliable. Here, I choose to neglect the latter possibility.

Deriving the detector design consists essentially of spelling out the functional implications of the demand on each cell to compute the test quantity eqn (2) in real time, given an appropriate choice of the signal densities $p_i(\rho_i; s_i)$. Thus, the general structure is always the same, a sum of terms depending on s_i, but the precise processing of these inputs to form the terms in the sum is dependent on the form of $p_i(\rho_i; s_i)$.

As a very coarse preview, Fig. 2 outlines the functional structure of the design. Both the general structure indicated here, and the detailed specification of various parts will be derived from the formal test eqn (2).

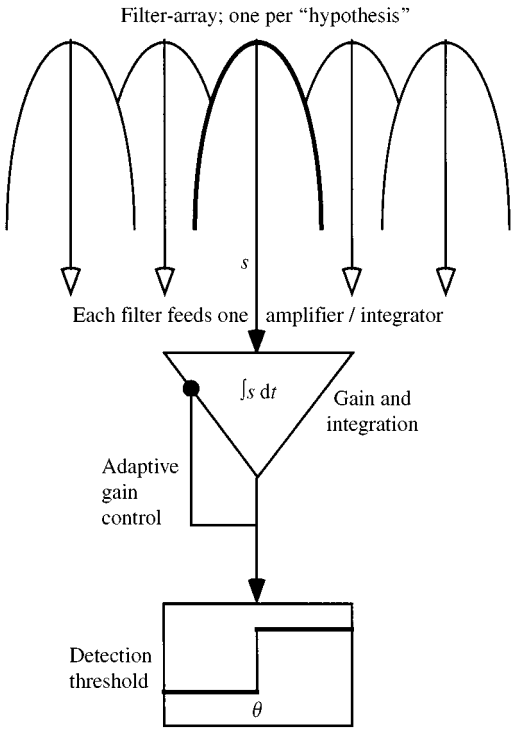

FIG. 2. The basic functional structure of how many detector units (T cells) should filter and integrate their available signals before detection. See the text for the derivation of this scheme from first principles, and for detailed functional specification of its parts.

2.2.1. *Multiple Hypothesis Testing ↦ Diverse Set of Selective Receptors*

Many unknown pathogen types can appear, and each can produce a few ligands out of a huge set of possible types. Dropping the temporary simplification of a single known intruder, the single Neyman–Pearson test generalizes to "multiple hypothesis" testing, with each hypothesis associated with the presence of one intruder-related ligand type. Thus, one needs to run many Neyman–Pearson tests in parallel, one for each hypothesis. To cover all possible ligands, a very large set of independent units (cells) is needed. The general form of their detection quantities is unchanged, so we may focus on any one of them, still specified by eqn (2).

Each of the available signals s_i should add a contribution $f_i(s_i) = \log p_i(\rho_i; s_i) - \log p_i(0; s_i)$ to the test quantity under consideration. With i indexing all ligands, the subtraction of terms in $f_i(s_i)$ implies that each detector must be equipped with a specific "filter" which nulls out all but a few of the available inputs s_i To see this most simply, consider the extreme case of signals unrelated to the truth of a certain hypothesis which is being tested. Thus, $p_i(\rho_i; s_i) = p_i(0; s_i)$ for these signals, and we get $f_i(s_i) = 0$, which means that the detector must stop such signals from contributing to its test quantity. This justifies the intuitive notion that "interference" must be filtered out early.

When considering the few remaining term(s) $f_i(s_i) > 0$, one notices their dependence on the (unknown) SNR ρ_i (see also Fig. 1). At any fixed value of ρ_i, each summand in eqn (2) determines a definite function $f_i(s_i)$. For simplicity, I assume until further notice that $f_i(s_i)$ increases monotonically with s_i, and that its ρ dependence can be absorbed into some (irrelevant) monotonicity—in statistical terms, eqn (2) is assumed to be a "uniformly most powerful" test with respect to ρ. This is likely to be true in reality, and it can indeed be shown for the examples worked out later (until stated otherwise).

Under these weak conditions, eqn (2) implies that the specific filter in front of a detector should essentially transmit only those signals which depend most strongly on the presence of the ligand in question. In immunologial terms: there should be many clones of cells, each having just one type of receptor, with enough specificity for avoiding significant interference from the other ligands which a cell is likely to encounter in its lifetime (the much larger number of *possible* ligands is of course irrelevant).

The minimal diversity of the set of receptor types should scale with the required degree of specificity, since coverage of all possible intruder-related ligands must be guaranteed. More quantitative conclusions about the required specificity and diversity can be derived by putting more detailed optimization demands on the system (deBoer & Perelson, 1993; Borghans & deBoer, 1998; Borghans *et al*., 1999).

In any case, the general feature of a large diversity of specific T cells is seen to follow from basic detection theory when the type of signal produced by an intruder is essentially unpredictable.

2.2.2. *Early Amplification and Integration* ↦ "*Serial TCR Triggering*"

The need to evaluate eqn (2) implies also that the signals which are selected specifically for each hypothesis must be amplified as soon as possible, and integrated over an appropriate time interval. Early-stage amplification is needed because ligands can occur in such low concentration that their signals s_i are low-rate Poisson processes, as generated by individual receptor binding-events. Mere signal transduction (essentially "copying") before using the signal to trigger some "large" response will not do: copying can only add more sampling-error, and thus the original s_i required by eqn (2) would be irretrievably degraded.

"Amplification" in this context implies that each available ligand molecule must generate a larger number of molecules. In other words, the ligand must act as a *catalyst*. This leaves no doubt about what substrate the catalyst should act on: it must be the specific receptor to which it binds. The simplest kinetic scheme is then

$$\mathrm{L} + \mathrm{T} \underset{k_d}{\overset{k_a}{\rightleftharpoons}} (\mathrm{LT}) \overset{k_c}{\rightarrow} (\mathrm{LU}) \overset{k_d}{\rightarrow} \mathrm{L} + \mathrm{U}. \quad (3)$$

Thus, ligand (L) and TCR (T) associate at rate k_a to a complex (LT); its TCR-part undergoes some irreversible conversion(s) to a product U, say, at a rate k_c. Dissociation of the complex, at a rate k_d (irrespective of the conversion-state of the TCR) then frees the ligand L. Thus, many T's can be converted into U's per ligand molecule. In fact, the typical time for an L molecule to go through a cycle of association and dissociation is $\tau_c = \tau_a + \tau_d = (k_a T)^{-1} + k_d^{-1}$. Per cycle, a T is converted to a U with probability $k_c/(k_c + k_d)$. The U's can be processed further to mount a cell-level response. Apart from the need to apply a threshold, eqn (2) does not specify such late stages, which are indeed no longer critical to the sensitivity of the system. Hence, I ignore the late stages here.

Note that even this simplest scheme not only amplifies, but also integrates the input signal. Integration is actually another of the features demanded by eqn (2), as follows: the "raw" input $s(t)$ is a series of spikes indicating individual ligand molecule binding events; thus, $s(t)$ is a Poisson process with rate $r = k_a LT$. The rate can be taken as a constant over a single integration time-interval τ, i.e. the contact time between a T cell and an APC (several hours). Dividing τ into slots of width δ, the numbers n_i of binding events in slots i are independent Poisson samples with $p(n_i) = (r\delta)^{n_i} \mathrm{e}^{-r\delta}/n_i!$ The generality of the Neyman–Pearson formalism eqn (2) implies that it must also apply to the numbers n_i, which then act as the formal "signals" s_i. Writing the background Poisson-rate as r_0, one obtains

$$\lambda = \sum_i n_i \ln(r/r_0) + (r_0 - r)\delta$$

$$\cong \sum_i n_i \equiv \int_0^\tau s(t)\,\mathrm{d}t. \quad (4)$$

With r being constant on the time-scale τ, one can drop the factors $\ln(r/r_0)$ and offsets $(r_0 - r)\delta$ from the sum, since this only constitutes another monotonic transformation on λ. Likewise, the choice of δ is irrelevant. The end result is temporal integration of $s(t)$.

As noted already, this is carried out "for free" by the same scheme which was implied by the demand for selective amplification. Thus, three

functionally distinct subtasks become interwoven in their biochemically simplest implementation. Evolution could favour solving several functionally connected subtasks "for the price of one", unlike modern engineering which prefers implementations to reflect the modularity of their functional design.

To quantify the basic behaviour of this scheme, it is sufficient for now to consider only expectation values of concentrations, and to work in the regime where the total amount of ligand L^* is small relative to the TCR level T, making the U response linear in L^*. To model large-L saturation, one would have to account not only for finite T, but also for possible saturation in the conversion step(s) which are now modelled simply by a fixed rate k_c. Such implementation aspects fall outside the scope of this paper, which focusses on the most basic functional aspects.

We are interested in the total output $U(\tau) - U(0)$ produced by L^*. In the L-linear regime, the performance is described by a net "gain" factor $\alpha = (U(\tau) - U(0))/L^*$, which reflects the combined effects of amplification and integration. Under a quasi steady-state approximation (slow T depletion, and $\tau \gg \tau_c$), we get $\alpha = (\tau/L^*)\partial_t U$ and $L^* = (1 + KT)L$; $K \equiv k_a/k_d$, with

$$\partial_t U = k_d[\mathrm{LU}] = k_c[\mathrm{LT}] = \frac{k_c k_a T L}{k_c + k_d}$$

$$= \left(\frac{k_c k_d}{k_c + k_d}\right)\left(\frac{KT}{1 + KT}\right)L^*. \qquad (5)$$

A contour plot of α appears in Fig. 3. To reduce the parameter set, without loss of generality, the plot actually shows a rescaled gain factor $\tilde{\alpha} = \alpha/(\tau k_c)$, equivalent to using a formal integration time $1/k_c$. Indeed, $\tilde{\alpha}$ is a function of only two rescaled rates $\tilde{k}_a = k_a/k_c$, and $\tilde{k}_d = k_d/k_c$, namely,

$$\tilde{\alpha} = \left(\frac{\tilde{k}_d}{1 + \tilde{k}_d}\right)\left(\frac{\tilde{k}_a T}{\tilde{k}_d + \tilde{k}_a T}\right). \qquad (6)$$

The plot also helps to delimit the parameter range where suitable behaviour occurs: one notes that the gain approaches a plateau $\tilde{\alpha} \to 1$ for parameters within the sector $\tilde{k}_a T > \tilde{k}_d > 1$. To operate in a plateau regime would spoil the antigen-specificity of the receptor, by wasting the gain differences between different ligands induced by the specific receptor, whose action can be represented as assigning pairs of rates (k_a, k_d) to any possible ligand. (In passing, note that the standard notion that receptors rank the ligands in terms of $K = k_a/k_d$ is inadequate for describing the function of this scheme.) Avoiding the $\tilde{\alpha} \to 1$ plateau in a robust way is possible because diffusion puts an upper limit on k_a. Indeed, this limit appears to be reached by many ligands "similar" to the best-matching one (Alam *et al.*, 1996). Simply keeping the TCR level T below a fixed maximum such that $\max(\tilde{k}_a T) \leqslant 1$, given some k_c, guarantees that the system stays below the tip of the forbidden sector for any ligand. Under this restriction, amplification maintains selectivity for ligands, based on their receptor-specific rate pairs (k_a, k_d).

Staying away from the maximum-gain plateau incurs some loss (roughly $\tilde{\alpha} \leqslant 1/6$), but this can be made up by the scale factor τk_c, which can be of order 10^3, given the observed $\tau = \mathcal{O}(10^4)$s and a plausible conversion rate $k_c = \mathcal{O}(0.1)\,\mathrm{s}^{-1}$ which is roughly of the order of the typical k_d for

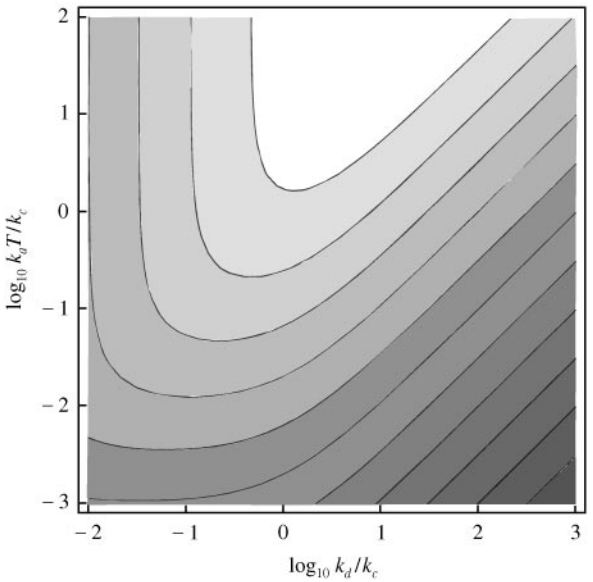

FIG. 3. Logarithmic intensity plot of the net gain factor $\tilde{\alpha}$ in the basic serial triggering scheme, eqn (3). Lighter grey means larger $\tilde{\alpha}$. Contours are drawn at $\tilde{\alpha} = 4^{-n}$, with $n = 1, \ldots, 11$. Note that a fixed affinity $K = k_a/k_d = \tilde{k}_a/\tilde{k}_d$, at constant TCR level T, corresponds to a line parallel to the diagonal $\tilde{k}_a T = \tilde{k}_d$.

"reasonable" ligands (Alam *et al.*, 1996). The resulting net gain $\alpha = \mathcal{O}(10^2)$ fits well with observations (Valitutti *et al.*, 1995) on the number of TCRs lost from the cell surface (down-regulation) per ligand molecule during the course of T cell activation. Indeed, continued TCR-conversion will deplete T, and U will eventually saturate to $U(\infty) = T(0)$. As long as the threshold θ is below about $T(0)/2$, the linear integration approximation used here is reasonable. Accumulation and thresholding of converted TCRs (U) has also been quantified experimentally ["TCR-counting" (Rothenberg, 1996; Viola & Lanzavecchia, 1996)].

Before completing this topic, it should be noted that adding another conversion step (Rabinowitz *et al.*, 1996) to the scheme, say (LU) → (LV) with the same rate k_c, makes $\tilde{k}_d$ independently available as the ratio U/V. As in the many-step scheme proposed before (McKeithan, 1995), this allows a T cell to *discriminate* between ligands which differ in their (k_a, k_d)-pair, even if they produce the same total signal ($U + V$), which still corresponds to lying on a contour in Fig. 3. The k_a-discrimination starts to fail for $\tilde{k}_a T > \tilde{k}_d$, where $\tilde{\alpha}$-contours converge to $\tilde{k}_d$-contours. Kinetic discrimination will not be analysed in more detail here, since the focus is on *detection*. For this, it suffices to have characterized the features of selective amplification and integration.

2.2.3. *Adaptive Background Scaling Controls $P_f \mapsto$ "Tolerance"*

To complete a first working design, we need to ask what is required for keeping the false alarm rate P_f of each detector very low. Specifying a precise value is hard, since not every false alarm will trigger a full autoimmune disease—it is likely that additional checks are done (e.g. by effectors) to reduce this risk. Nevertheless, the probability to induce disease will rise with P_f, since the checks will also have some failure rate. As a rough guess, P_f should scale inversely with the number of lymphocytes.

As already illustrated in Fig. 1, a low P_f requires the detection threshold to lie far into the upper tail of the background probability density of its test quantity u_i, which is the number of type-i TCRs converted in time τ. Writing the densities as $p_i(\rho_i; u_i)$, parameterized by the specific SNR ρ_i, the false alarm rates are $P_f = \int_\theta^\infty p_i(0; u_i)\,\mathrm{d}u_i$. (If θ were to be i-dependent, one simply redefines u_i in units of θ_i.) Clearly, the system must compensate somehow for the i-dependence of the densities $p_i(0; u_i)$, i.e. each cell must adapt its detection process to the specific background that it samples. In an early study addressing this problem, Grossman & Paul (1992) proposed a dynamical model of how T cells could adjust θ to "track" their background level. Here, within the context of optimal design, an adaptive "gain-control" scheme emerges if one simply continues the interpretation of the formal design in terms of TCR-triggering.

The starting point is that different receptor types i allow different ligands to dominate their u_i values, and these ligands will occur with widely different mean levels L^* and kinetic parameters k_a and k_d. Directly or via the gain α, these differences determine an i-dependent scale for u_i. Hence, one expects that, to a good approximation, all $p_i(0; u_i)$ are scaled copies of each other. To formalize this, consider a $\hat{u}_i$ which would be obtained if all other relevant parameters (T, τ, k_c) were i-independent. Its characteristic scale due to different dominant ligands is then measured most simply by the expected value $b_i = \int_0^\infty \hat{u}_i p_i(0; \hat{u}_i)\,\mathrm{d}\hat{u}_i$. The scaling assumption means that $p_i(0; \hat{u}_i) = b_i^{-1}\, p(0; \hat{u}_i/b_i)$, with $p(0; \hat{u}_i/b_i)$ being the i-independent density of the *relative* fluctuations due to T cells finding their ligands presented by various types of APC, at various times and places. It should also cover the Poisson-noise remaining after integration.

The functional implications are clear: the test quantity to use is $u_i = \hat{u}_i/b_i$. With this rescaling, setting θ to achieve some allowed P_f needs to be done only once, and for all clone types i, e.g. during cell maturation. Scaling $\hat{u}_i$ by b_i also fits well into the scheme which was already derived. The simplest approach is to scale the gains α_i by b_i. Inspecting eqn (6), one notes that the TCR-level T is the natural parameter to be used as the required "gain-control". Indeed, in the proper regime (see Fig. 3), α depends quasi-linearly on T, as previously determined.

What remains then is the need to set up the type-i TCR levels $T_i \propto 1/b_i$. This too fits well into the existing scheme. In fact, the required result arises naturally by down-regulation of T_i during

prolonged exposure to the background. This may occur in an "adaptation" phase early in the life of the cell, or even throughout its active life. To be explicit, say that new T's are created at a rate σ while the cells sample background, preferably with response disabled. Neglecting fluctuations, T_i will approach a steady state T_i^*, i.e. the influx σ equals the efflux, which is $\partial_t U_i \propto b_i T_i^*$ as analysed before. Thus, one gets $T_i^* \propto \sigma/b_i$, as required.

Additional means of gain-control clearly exist, e.g. the cell could adapt its down-stream gain (phosphorylation cascade), its contact time τ, or its threshold θ (Grossman & Paul, 1992). The net effects would be similar to adapting the TCR-level T, and combining several such mechanisms would increase the range over which the cell can control its overall sensitivity.

The effect of an adaptive background rescaling constitutes a form of "tolerization"—it guarantees that potentially self-reactive T cells have the same very low P_f as T cells which encounter virtually no matching ligands in the background. In addition, the self-reactive T cells would require very effective stimuli to get activated. Given more moderate stimuli, they would appear to be "anergic". When the adaptation is allowed to occur continually, but on a time-scale much slower than τ, the result would be a form of "peripheral tolerance". As an extension of the scheme for initial adaptation, one could simply delete cells whose T_i after adaptation would be so low that their u_i could never reach θ under physiological conditions. This approach can be identified with thymic "negative selection".

Note that in such schemes, TCR down-regulation, anergy and negative selection are predicted to be driven by ligands which act as *agonists* peripherally. Qualitatively, this fits classical observations, but the present scheme regulates cells in a gradual manner, which has the advantage that no true "holes in the repertoire" can form. Instead, each clone is calibrated to maintain a fixed low false-alarm rate P_f while maximizing P_d.

3. Detection Performance of the Basic Design

3.1. AN IDEALIZED SIMPLE CASE: FIXED INTEGRATION TIME

To gain a first impression of detection performance, and to set the stage for more realistic models to be analysed later, consider the limiting case in which the contact time τ between T cell and APC is precisely fixed. In this case, time units may be rescaled so that $\tau = 1$. We may also drop the specificity index i because each detection occurs independently of all others, and because the adaptive gain-control equalizes the background statistics of u_i.

Quantifying detection performance means determining the relation between detection rate P_d, SNR ρ and false alarm rate P_f. The function which determines this relation is the density $p(\rho; u)$ of the test quantity, since both P_d and P_f are defined as upper-tail integrals of this function. To compute $p(\rho; u)$, we need to take account of several sources of fluctuation which contribute to u.

First, recall that u is conditionally Poisson-distributed, i.e. given an expected value h of u, one has $p(u|h) = h^u e^{-h}/u!$. Per detection attempt, h should be a weighted sum over ligand levels, with weights given by the ligand-specific gain factors α, depending on various kinetic and other parameters. However, there is no use in writing this out here; what matters is simply to find the appropriate h-density $p(\rho; h)$.

Indeed, h must also fluctuate, since T cells sample ligands at various times and places, via various APCs. Both background and intruder-related ligands contribute to h, and each will fluctuate. I assume the two terms to be additive and independent, and thus the h-density $p(\rho; h)$ for general SNR ρ is simply the convolution of the densities of the two terms. These densities, which are experimentally unknown so far, can be chosen without adding *ad hoc* assumptions, as follows.

Because of the adaptive gain-control, the background-term density $p(0; h)$ has some fixed mean, say m. Being generated by a similar random process, the purely intruder-derived term must have a density which differs from $p(0; h)$ only by a scaling. The scale factor then has to be the SNR-parameter ρ. To compute illustrative examples, one now needs to specify an explicit $p(0; h)$. Given that h is nonnegative and that it has mean value m in this case, the unique solution $p(0; h) = m^{-1} e^{-h/m}$ is determined by parsimony (formalized as maximum entropy; Shannon,

1948). By scaling, the density of the intruder term becomes $(m\rho)^{-1}e^{-h/(m\rho)}$.

Exponential densities also seem not unreasonable biologically: values from zero up to about the mean are roughly equally likely, creating a substantial coefficient of variation, but much larger excursions than the mean quickly become unlikely. Of course, one would like to know the true *in vivo* density, but the present choice suffices for computing some illustrative examples. The general method of computation could be used with other $p(0; h)$, although perhaps only numerically instead of analytically.

By convolution, we obtain the density $p(\rho; h)$ for general SNR ρ as

$$p(\rho; h) = \frac{e^{-h/m}}{m} * \frac{e^{-h/(m\rho)}}{m\rho}$$

$$= \frac{e^{-h/(m\rho)} - e^{-h/m}}{m(\rho - 1)}. \quad (7)$$

To find the unconditional density $p(\rho; u)$ one integrates h out of the conditional Poisson density $p(u|h)$, i.e. $p(\rho; u) = \int_0^\infty p(u|h)p(\rho; h)\,dh$. This yields

$$p(\rho; u) = \frac{1}{u!m(\rho - 1)} \int_0^\infty h^u (e^{-h(m\rho + 1)/m\rho} - e^{-h(m+1)/m})\,dh$$

$$= \frac{1}{m(\rho - 1)} \times \left[\left(\frac{m\rho}{m\rho + 1}\right)^{u+1} - \left(\frac{m}{m+1}\right)^{u+1}\right]. \quad (8)$$

Having introduced h-level fluctuations, we need to check whether u still fits the Neyman–Pearson formalism, eqn (2), i.e. whether u is monotonically related to λ with a formal signal $s = u$. This can be shown to be the case. Thus, detection is declared when $u \geqslant \theta$, and thus the important quantities P_d and P_f are equal to the upper tail probabilities $Q(\rho; \theta) = \sum_{u=\theta}^\infty p(\rho; u)$, with either $\rho > 0$ or $\rho = 0$, respectively. Summing the two geometric series yields

$$Q(\rho; \theta) = \frac{1}{\rho - 1}\left[\rho\left(\frac{m\rho}{m\rho + 1}\right)^\theta - \left(\frac{m}{m+1}\right)^\theta\right] \quad (9)$$

$$\approx \frac{\rho e^{-\theta/(m\rho)} - e^{-\theta/m}}{\rho - 1}, \quad (10)$$

where the approximation is accurate throughout the relevant regime $\theta \gg 1$ as long as $m \gg 1$. Note that the approximate result, eqn (10), is equal to the upper tail probability $\int_\theta^\infty p(\rho; h)\,dh$ for the Poisson-mean h. This confirms the intuition that the discreteness and Poisson fluctuations of u become irrelevant when crossing the threshold requires many converted TCRs, as occurs in real T cells (Viola & Lanzavecchia, 1996). Thus, I proceed with the $m \gg 1$ approximation, which offers the analytical convenience of acting as if the detection uses h instead of u as test quantity. Since h is continuous, one may then absorb m into redefined units of h, so that the result eqn (10) for $Q(\rho; \theta)$ applies formally with $m \leftarrow 1$.

The required θ is found by solving $P_f \equiv Q(0; \theta)$. With $Q(0; \theta) = e^{-\theta}$, one gets $\theta = \ln P_f$, so that any practical false-alarm probability requires $\theta = 15$–30 (now in units of m, the mean number of TCRs triggered by background). The very slow dependence of θ on P_d is expected to be a robust feature, unless when the background has such a slowly decaying upper tail that its expectation diverges.

The detection rate $P_d = Q(\rho; \theta)$ of the most basic design is now fully determined as a function of SNR ρ, for any given P_f. Some P_d vs. ρ graphs are shown in Fig. 4, for $P_f = 10^{-5}$, 10^{-8} and 10^{-11}. The closeness of the curves confirms that changing P_f by orders of magnitude, which changes θ only weakly, has only a moderate effect on the "detection limit", which will be defined as the signal/noise ratio ρ at which a reasonably large P_d-criterion is reached.

3.2. CONTACT TIME FLUCTUATIONS CAN SPOIL SENSITIVITY

It is reasonable to ask whether the detection performance would be affected strongly by more realistic conditions. Indeed, one expects τ to

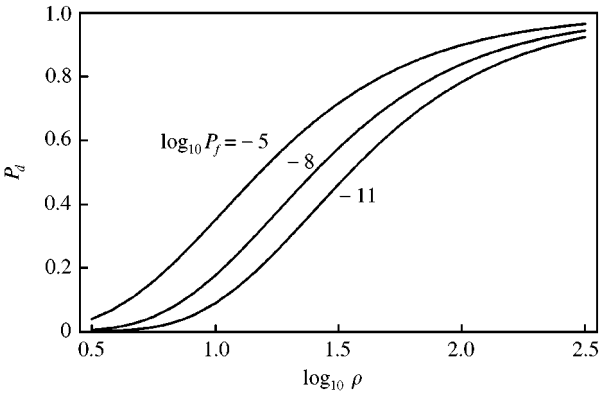

FIG. 4. Detection probability P_d vs. SNR ρ, at (left to right) $P_f = 10^{-5}$, 10^{-8} and 10^{-11}, for the simplest model, having a fixed contact-time between T cell and APC. Note that large changes in the allowed false-alarm rate P_f result in only rather small changes in the "detection limit", i.e., the signal/noise ratio ρ at which a reasonably large fixed P_d-level is reached.

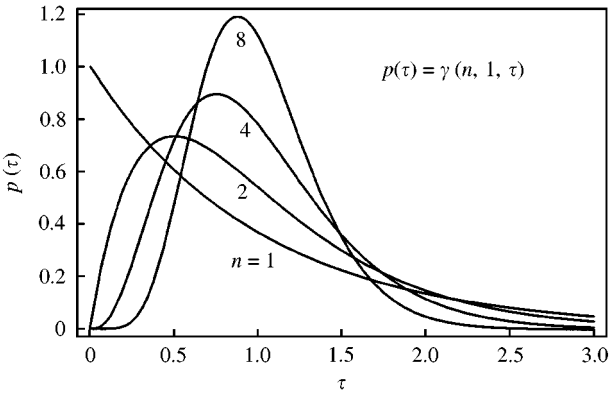

FIG. 5. The probability density of the (sealed) contact time τ, modelled as the unit-mean γ-density of order $n = 1, 2, 4, 8$. This model is exact when τ is controlled by n-th-order irreversible kinetics. The standard deviation is $1/\sqrt{n}$. The simple case of a fixed $\tau = 1$ thus corresponds to the limit $n \to \infty$.

fluctuate from one T cell activation attempt to another, and this would contribute multiplicative noise to the test quantity. The probability density $p(\tau)$ must be exponential when bound T cell/APC pairs dissociate with first-order kinetics, e.g. via one rate-limiting step. More generally, one may consider an n-step process, causing τ to become gamma-distributed with index n. The mean of τ may be scaled to one, yielding

$$p(\tau) = \gamma(n, 1; \tau) = n^n \tau^{n-1} e^{-n\tau}/\Gamma(n). \qquad (11)$$

This is plotted for several n in Fig. 5. Note that the relative τ-fluctuations decrease as $1/\sqrt{n}$. Real T cells may control τ in many ways, not to be determined *a priori*, but gamma-densities seem capable of modelling the net effect.

At a fixed τ, the (Poisson-mean) test quantity h has a conditional density $p(\rho; h|\tau)$ analogous to eqn (7),

$$p(\rho; h|\tau) = \frac{e^{-h/(\tau\rho)} - e^{-h/\tau}}{\tau(\rho - 1)} \qquad (12)$$

and the unconditional density $p(\rho; h)$ needed for computing P_d and P_f (neglecting discreteness and Poisson noise) follows by integrating out τ as $p(\rho; h) = \int_0^\infty p(\rho; h|\tau) p(\tau)\, d\tau$.

It is useful first to do this integral only for the background density $p(0; h|\tau) = \tau^{-1} e^{-h/\tau}$. This yields

$$p(0; h) = \frac{n^n}{\Gamma(n)} \int_0^\infty \tau^{n-2} \exp[-(n\tau + h/\tau)]\, d\tau$$

$$= \frac{2n}{\Gamma(n)} (nh)^{(n-1)/2} K_{n-1}[2\sqrt{nh}], \qquad (13)$$

where $K_n[x]$ is a well-known modified Bessel function (Abramowitz & Stegun, 1972). The upper tail probability $Q(0; \theta)$ is then

$$Q(0; \theta) = \int_0^\infty p(0; h)\, dh$$

$$= \frac{2}{\Gamma(n)} (n\theta)^{n/2} K_n[2\sqrt{n\theta}]. \qquad (14)$$

As before, θ follows by solving $P_f = Q(0; \theta)$. For realistic, very small P_f one may use known asymptotics (Abramowitz & Stegun, 1972) of $K_n[x]$ to approximate

$$P_f \approx \frac{\sqrt{\pi}}{\Gamma(n)} (n\theta)^{n/2 - 1/4} \exp(-2\sqrt{n\theta}), \qquad (15)$$

which also illustrates that the background now has a slower decaying tail than the $e^{-\theta}$ tail of the idealized case with $\tau = 1$, which is recovered for $n \to \infty$.

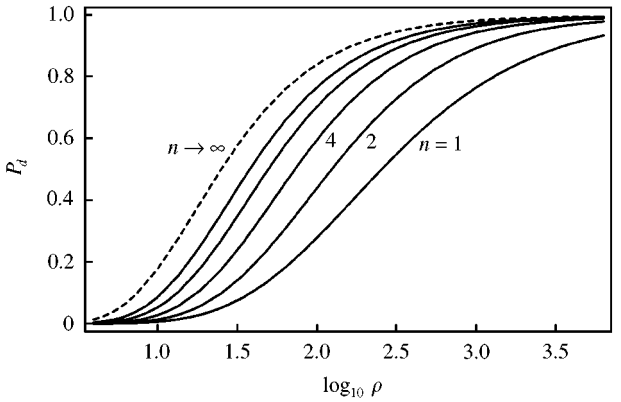

FIG. 6. P_d vs. SNR ρ, at $P_f = 10^{-8}$, for a stochastic contact time τ, controlled by an n-step process with $n = 1, 2, 4, 8, 16$. The (– – – –) ($n \to \infty$) graph is for fixed $\tau = 1$.

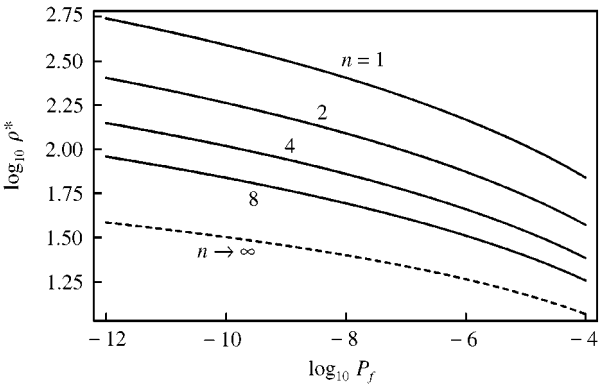

FIG. 7. The detection limit, i.e. the SNR ρ^* at which $P_d = 0.5$, plotted against the allowed false-alarm rate P_f. Drawn graphs are for stochastic τ, controlled by an n-step process with $n = 1, 2, 4, 8$. The (– – – –) ($n \to \infty$) graph is for fixed $\tau = 1$.

The upper tail probability $Q(\rho; \theta)$ for general SNR ρ is now easily expressed in terms of $Q(0; \theta)$, because $p(\rho; h|\tau)$ is related to $p(0; h|\tau)$ by linearity and scaling [see eqn (12)]. Thus, one can directly write down the result as

$$Q(\rho; \theta) = \frac{\rho Q(0; \theta/\rho) - Q(0; \theta)}{\rho - 1}. \qquad (16)$$

Representative graphs of $P_d = Q(\rho; \theta)$ are plotted in Fig. 6. Note the roughly ten-fold rise in detection limit when τ fluctuates exponentially ($n = 1$), relative to the $\tau = 1$ ideal. Even moderate control of τ fluctuations, as e.g. by n-step kinetics, earns a good improvement, but this saturates quickly for about $n > 4$.

In Fig. 7, I show how very large variations in P_f, which induce only moderate variations in θ, affect the "detection limit", defined as the SNR ρ^* at which detection occurs with a reasonably large P_d-criterion (P_d^*). For all examples, I take $P_d^* = 0.5$. Clearly, the overall chance of failing to detect an intruder must be much smaller than the choice $1 - P_d^* = 0.5$—perhaps 10^{-3} is a more realistic upper bound—but at $P_d = 0.5$, one needs only ten independent detection "attempts" to achieve this goal. Indeed, it is not unreasonable to assume that at least ten T cells per clone exist, or that any T cell could undergo sequential contacts with APCs which present the intruder ligand. The fact that with such multiple attempts, the total false alarm probability also accumulates (but only linearly!) should thus be incorporated into the choice of a very low P_f. This increases the detection limit only very weakly, as seen in Fig. 7. Note that for ill-controlled contact time (small n), the detection limit is slightly more dependent on P_f. This reflects the longer upper tail of the background density for small n, which makes θ more dependent on P_f.

4. Extending the Basic Design

4.1. WHAT TO DO WITH NON-SPECIFIC SIGNALS: CO-STIMULATION

The basic version of the design has exploited signals which are specific in the sense that each signal is required to contribute to only one of the many hypotheses being tested. However, it is to be expected that not all signals have such specificity. This prompts the question how (if at all) the design should use such signals. As before, the Neyman–Pearson formalism dictates the answer. Even before deriving it, we know that the performance of the extended system cannot be worse than the one using only specific signals: at worst, the formalism would yield the old system, with any non-specific signals dropped out of the optimal test quantity.

As a simple example, consider the extreme case of a single signal a which is fully non-specific, while the previously used signals $\{s_i\}$ are now explicitly assumed to be fully specific. This assumption sets up a useful caricature of the

immunological situation, in which a moderate number of more or less non-specific signals tend to occur. Note that the relevant meaning of a signal being "specific" is independent of its being generated by just one type of pathogen or by many. Thus, the same material, say LPS (which occurs in many bacteria), may even play both roles: it is a "specific" signal when it triggers an anti-LPS clonal response, and a "non-specific" signal when it somehow influences other responses.

The optimal way of using non-specific signal a is found as usual, by substituting the signal density into the likelihood ratio for a multiple hypothesis Neyman–Pearson test. The test quantity for hypothesis j will depend on a if the density $p(a)$ depends on the SNR parameter ρ_j. With fully non-specific a, this applies for all j, proving the need to use a. The fluctuations of a at any fixed set of ρ_j are likely to be essentially independent of those of the s_i, so we can still use the log-likelihood form of eqn (2). Simplifying to the case of rare infections, i.e. if at most a single ρ_j is positive in any time-interval of interest, the formal test quantities become

$$\lambda_j = \log p(\rho_j; a) - \log p(0; a) + \sum_i \log p_i(\rho_j; s_i) - \log p_i(0; s_i). \qquad (17)$$

Since the s_i-terms are unchanged, we may approximate them as before with their respective (Poisson-means) h_j. Analogous processing demanded by the new a-terms leads to a new test quantity $H_j = f(a) + h_j$, where $f(a)$ will be monotonically increasing in a for any reasonable $p(\rho_j; a)$. Thus, we should simply add the non-specific signal [processed as $f(a)$] to the previously specified specific signal terms. Equivalently, one could change the previous threshold θ by subtracting $f(a)$.

To compute how much improvement in sensitivity is obtained by this use of extra information, one would need to specify $p(\rho_j; a)$ and generalize the previous performance analysis. Although this is a viable project, it exceeds the scope of the present paper, and I therefore leave this analysis for another occasion.

For now, it is sufficient to have established the principle that optimal detection should use not just the specific signal h_j, but also any non-specific signals related to the presence of intruders in general. Immunologically, the fact that the $f(a)$ term does not depend on j suggests that the most logical source of a would be the APCs, and so we may tentatively identify a with co-stimulation, e.g. "signal-2", by which activated APC signal their state via B7-type ligands to CD28 receptors on T cells. APCs are already charged with the task of presenting the specific s_i ("signal-1") in a broadly unbiased way, so that they could quite simply compute a from the s_i or other sources, such as epitopes which occur in many types of pathogen (e.g. LPS), or biochemical evidence that the APC itself is infected. The functional effect of co-stimulation on T cells also fits what the design demands: it lowers the T cell threshold (Viola & Lanzavecchia, 1996) and probably boosts the signal amplification by recruiting phosphorylation machinery (Wülfing & Davis, 1998; Viola *et al.*, 1999). Both effects fit the design result, given the freedom in choosing some fixed monotonicity.

4.2. CORRECTION FOR EARLY SIGNAL BLURRING: ANTAGONISM

Before signals are actually detected, the distinction between "similar" but independent signals will unavoidably be blurred to some extent, thereby creating spurious mutual correlations which would, when left uncorrected, reduce the information accessible to the detectors and thus affect their performance. Blurring occurs in any multi-stage signal processing system. In our case, the many steps involved in antigen pre-processing and presentation will cause a *nominally* unique antigen to appear to T cells as random samples from a set of similar but non-identical signal types. Even with a single epitope, protein processing may excise ligands with slightly varying lengths; or a nominally unique ligand may be presented in slightly varying modified forms, e.g. by co-binding to MHC and TCR in distinct conformations, which also fluctuate by thermal random motion. When many similar but unrelated primary signals are sent through such a processing chain, their signals become mixed, even when each TCR is nominally very specific. I quantify this process and study the consequences of its (partial) remedy.

To ease analysing the problem, one may associate each TCR-type with a sampling point in some high-dimensional "shape-space" (Segel & Perelson, 1988). When the sampling is denser than the typical distance over which signals become blurred by preprocessing, nearby TCRs will send correlated signals to their detectors. Information is then lost: the fine structure of the signal distribution across shape-space is blurred, which limits implementing the demand of eqn (2) to separately amplify any initially independent signals before threshold detection. The most obvious conclusion is that the pre-processing should try to minimize the blurring it causes, but physical and biochemical limitations prevent reducing blur to zero. Thus, one is forced to try getting rid of blurring at the level of T cells. Within bounds, this is possible, as follows.

The aim is to find a transformation of the signals to which T cells have direct access, say the binding rates $r(x)$, such that it reduces the *overall* blurring which occurs between the (not directly accessible) primary signals, denoted by $s(x)$, and the set of (local mean) test quantities $h(x)$. Note that the discrete ligand index i has been replaced by a continuous D-dimensional coordinate x in shape space. As expected (and shown below), it is impossible to fully undo the blurring. However, the scale of the overall blurring can be reduced by a moderate factor, and this is already a great bonus in a high-D space, where the number of effectively independent detections (i.e. the proper functional measure of diversity) scales inversely with σ^D, the volume of a cube with the overall blur scale σ as its side.

The required decorrelation approach is well known (as "pre-whitening") in radar and communication engineering. Here, I present a simple example to illustrate the essential approach and its results.

4.2.1. *Partial Decorrelation ↦ Antagonistic Centre-Surround Response*

Under linearity, blurring is described by convolution. Thus, one has $r(x) = B(x) * s(x)$, with a blur kernel $B(x)$ due to "sloppy" preprocessing. For the sake of computing a simple explicit example, I take $B(x)$ as being exponential and separable, as seems reasonable. Thus,

$$B(x) = \prod_{j=1}^{D} B_1(x_j) = \exp\left(- \sum_{j=1}^{D} |x_j| \right). \quad (18)$$

Fourier transformation, denoted by adding a tilde to variables, e.g. $\tilde{r}(q) \equiv \mathscr{F}[r(x)](q)$ where q is the spectral frequency, is useful since it converts convolution into spectral multiplication. Thus, we have $\tilde{r}(q) = \tilde{B}(q)\tilde{s}(q)$, with

$$\tilde{B}(q) \equiv \mathscr{F}[B(x)](q) = \prod_{j=1}^{D} \mathscr{F}[B_1(x_j)](q_j)$$

$$= \prod_{j=1}^{D} (1 + q_j^2)^{-1}. \quad (19)$$

Similarly, writing the effect of TCR-selectivity and any subsequent linear processing for undoing the blurring as $u(x) = C(x) * r(x)$, we have $\tilde{u}(q) = \tilde{C}(q)\tilde{B}(q)\tilde{s}(q)$. Full de-blurring would require $\tilde{C}(q) = 1/\tilde{B}(q)$, but inspection of eqn (19) reveals that this is actually an illposed (structurally unstable) problem: the spectral amplitude of the full inverse $1/\tilde{B}(q)$ is seen to diverge for $|q| \to \infty$. However, we can still define a "regularized" pseudo-inverse: the simplest choice is to multiply the "pure" inverse by a Gaussian $\exp(-\sigma^2|q|^2/2)$ to damp the high-$|q|$ amplitude. Thus, we obtain the appropriate T cell kernel $C(x)$ as

$$\tilde{C}(q) = \prod_{j=1}^{D} (1 + q_j^2) \exp(-\sigma^2 q_j^2/2) \quad (20)$$

$$\Rightarrow C(x) = (\sqrt{2\pi\sigma})^{-n} \prod_{j=1}^{D} \left(1 - \frac{\partial^2}{\partial x_j^2}\right) \exp\left(-\frac{x_j^2}{2\sigma^2}\right)$$

$$= (\sqrt{2\pi\sigma})^{-n} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

$$\times \prod_{j=1}^{D} \left(1 + \frac{1 - (x_j/\sigma)^2}{\sigma^2}\right). \quad (21)$$

For any choice of σ, the $C(x)$ kernel has a "mexican hat" shape, positive near the optimal ligand, negative ("antagonistic") for ligands at distances of a few times σ, and quickly becoming negligible for more distant ligands. A practical example with $\sigma = 1/2$ is shown in Fig. 8.

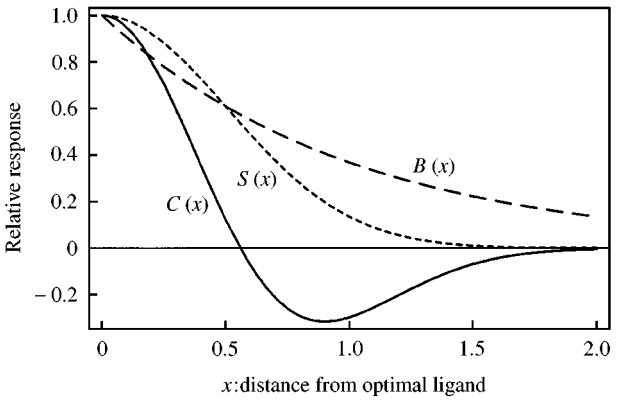

FIG. 8. Blurring and deblurring of the response to signals, at three stages of antigen processing. All graphs show blur-kernels, scaled to unit-maximum, sectioned along one of the shape-space dimensions (see text). $B(x)$: blurring due to antigen pre-processing and presentation. $C(x)$: Deblurring kernel (with $\sigma = 0.5$), showing antagonism for ligands in a halo around the TCR's best match. $S(x)$: overall result, being the convolution of $B(x)$ and $C(x)$. Note the much smaller tails than those of $B(x)$.

The overall selectivity $S(x)$, incorporating all steps from pre-processing through to T cell activation, is then given by $S(x) \equiv B(x) * C(x) = (\sqrt{2\pi}\sigma)^{-n} \exp[-|x|^2/(2\sigma^2)]$. Thus, we obtain a Gaussian "specificity" at the system level, with its width σ corresponding to the parameter in the regularisation of the ill-posed inverse $C(x)$. Attempting to reduce to $\sigma \ll 1$ would reveal the underlying structural instability: the $C(x)$ kernel then gets ever larger negative tails, meaning that r becomes a small difference between two ever larger terms. Another lower bound to σ is given by the typical distance between optimal ligands for neighbouring receptors. Smaller σ would lead to "holes" in the coverage of the space of ligand shapes.

It may be worth noting that the designed antagonism could only be visible experimentally when probing with slightly mutated ligands. Mutations at the level of unprocessed pathogens should not "see" the antagonism, at least on average, since the overall effect of blurred processing and detector antagonism is designed to produce a single, short-tailed peak $S(x)$ in shape space, as shown in Fig. 8. Pathogen mutants could only exploit imperfections in the implementation, or attempt to sabotage its machinery.

In reality, antagonistic T cell responses are indeed often found when using single-mutant forms of (presumably) optimal ligands (Sloan-Lancaster & Allan, 1996). So far, the function of this antagonism has been puzzling; in fact, it has been viewed as a dangerous flaw ("Achilles heel", Alam *et al.*, 1996) of T cells, which might allow mutant pathogens an easy escape from detection. The present derivation provides a clear functional role for antagonism, and denies the putative danger as long as the system fits the specifications derived here.

Although the design allows several implementation schemes, it seems natural to attribute the implementation of antagonism to the multi-step phosphorylation of triggered TCRs (Neumeister Kersh *et al.*, 1998), as predicted by the "kinetic proofreading" model (McKeithan, 1995). Earlier in this paper, I noted how the two-step simplification of this approach (Rabinowitz *et al.*, 1996) emerges as the natural extension of the design for selective signal amplification and integration. A two-step scheme can indeed be functionally as powerful as any multi-step scheme since ligand binding has just two parameters, k_a, k_d.

4.2.2. *Adaptation Revisited*: *Positive Selection on Self-Peptides*

Independent of the implementation, another prediction follows if one allows the deblurring scheme to be already in effect when the T cell gain is adaptively calibrated by its background signals. Consider T cells which do not find their cognate ligand among the background during thymic maturation, and assume that the pre-processing and presentation of self-antigen in thymus is less blurred than in the periphery. If the T cells do see ligands which lie in the halo around the perfect match, their gain-control T will *increase*. It would be natural to use this as a "survival" signal, consistent with the earlier assumption that strong T depletion would lead to cell death. This fits at least qualitatively with thymic "positive selection". The negative selection process noted earlier would continue, driven by ligands with a near-perfect fit, lying in the central positive lobe of $C(x)$.

Before considering the effect of a deblurring operator, the adaptation of cells to their

background could only decrease their gain factor (most easily by means of TCR-level T). The new scheme predicts that T cells could also be *positively* selected by self-ligands (Barton & Rudensky, 1999). The apparent paradox of two opposite effects being due to the same set of ligands is now resolved by pointing to the role of the mismatch (distance x) between the cognate and the actual ligands. Moreover, the design shows how the existence of such processes emerges naturally from the task of handling blurred signals.

4.3. SPIKY BACKGROUND ↦ HIGH-ZONE TOLERANCE

Throughout the design process discussed so far, the optimal test quantity was found to be a non-decreasing function of the input, say of the Poisson rates r of ligand-binding events. This monotonicity is indeed found for a very wide class of background and signal densities $p(\rho; r)$. However, it is useful to explore how robust this feature is. In particular, it is relevant to ask when the optimal test quantity can be a one-humped function of r, implying that actual detection can occur only in a finite interval of input signal strengths. As shown below, this type of response is in fact required when the background is more "spiky" than hitherto assumed, i.e. when its spatial and temporal fluctuations contain rare but very large peaks. A reduced or absent response for very strong stimulation is in fact a classical feature of real T cells (Matis *et al.*, 1983; Mitchison, 1964) known as "high-zone-tolerance". Here, I do not explore what may be a whole range of distinct processes to which this term is often applied—my aim is merely to show how the demand for a "humped" type of response is logically implied by optimal detection in a spiky background.

The only extension of previous assumptions is to add to the usual background a low-probability ε of spikes with mean strength $\zeta \gg 1$. Thus, after the usual early steps of signal processing we have (neglecting Poisson-noise)

$$p(0; h) = (1 - \varepsilon)\mathrm{e}^{-h} + \frac{\varepsilon}{\zeta}\mathrm{e}^{-h/\zeta}. \qquad (22)$$

The qualitative change from the design so far will be that the filtered and integrated signal h is now no longer monotonically related to the test quantity H required for optimal detection, which will be derived below.

Indeed, if one simply used h as test quantity, very bad performance would occur as soon as $\varepsilon > P_f$, i.e. whenever the spikes are not extremely sparse: the detection limit ρ^* is roughly proportional to the threshold θ, and this would then have to be raised to a few times ζ to maintain a low P_f. Since ζ is easily a few orders of magnitude above the unit-scale of the non-spiky part of the background, the system would essentially become "blinded" by the spikes.

The required strategy follows by applying the usual formalism, yielding the correct test quantity $H = \log[p(\rho; h)/p(0; h)]$, where now the h-densities are given by

$$\begin{aligned} p(\rho; h) &= p(0; h) * \frac{1}{\rho}\mathrm{e}^{-h/\rho} \\ &= \frac{1 - \varepsilon}{\rho - 1}(\mathrm{e}^{-h/\rho} - \mathrm{e}^{-h}) \\ &\quad + \frac{\varepsilon}{\rho - \zeta}(\mathrm{e}^{-h/\rho} - \mathrm{e}^{-h/\zeta}). \end{aligned} \qquad (23)$$

As usual, the result still depends on parameters such as the SNR ρ, but H remains a "one-humped" function of h for all relevant parameter values. This is illustrated in Fig. 9, and can be supported by formal analysis.

Setting θ even slightly above the high-h asymptote $H = 0$ prevents most of the background spikes from causing a false alarm. In terms of h (or its monotonic equivalents), the conclusion is that *two* thresholds are required, say a lower one θ_- and an upper one θ_+. The ρ-dependence of H requires a slightly modified approach to cover the most relevant range of ρ. The simplest modification is to fix the two h-thresholds independently, instead of deriving them from one θ acting on the formal H. The ε and ζ values do not matter much, as long as they are realistically small, respectively large.

To reduce the false alarms on background spikes from an often unacceptable "raw" rate near ε to the normal P_f, the θ_+ value must be set to about $P_f\zeta/\varepsilon$, which can easily be of order 10^3.

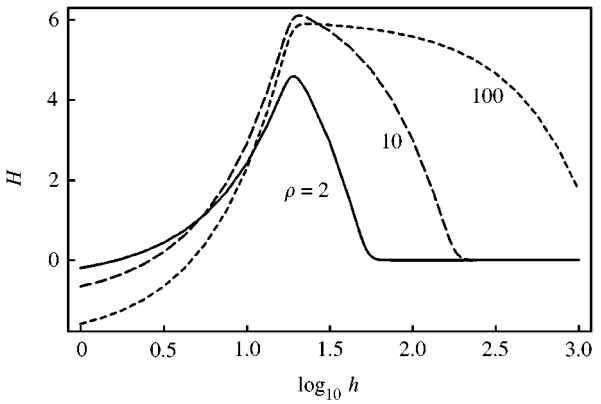

FIG. 9. With spiky background, the optimal test quantity H becomes a non-monotonic function of the linearly integrated ligand signal h. The shape of $H(h)$ depends also on e.g. the SNR ρ (three graphs with $\rho = 2, 10, 100$), but it remains a one-humped function for almost any parameter values of practical importance. In immunological terms, the theory demands "high-zone tolerance".

For the scheme to work, the probability that θ_+ is exceeded when an intruder ligand is present should stay below, say, 0.1. In view of the detection limit in non-spiky background (very roughly $\rho \approx 100$), this condition presents no problem when intruders just become detectable. At very high ρ, the signal might exceed θ_+ and become undetectable again. This may appear to be a nasty side-effect, but this regime would never be reached in practice, where the pathogen level starts at a small value, and grows until it triggers an immune response. Thus, if an effective response can be triggered at all, it will control the pathogen soon after crossing θ_-, and prevent it from reaching θ_+.

A quantitative evaluation of performance with optimized θ_+ and θ_- is possible, but this would lead too far beyond the present scope. For now, it is enough to have established the general conclusion that optimal detection of signals in spiky background requires "high-zone tolerance" in one form or another. Insofar as this occurs in real T cells, we have identified a definite function for it.

4.4. ANTIGEN-SPECIFIC ACTIVATION OF NON-SPECIFIC RESPONSES

So far, the design has been based on a task which includes the assumption that each type of intruder, when detected, requires a response which is directed specifically to it (or in fact, to its specific epitopes). This assumption is directly responsible for the general design feature of parallel testing of a huge set of hypotheses, implemented by many clones of T cells which can trigger responses of the same specificity as they use for detection. Without dropping this assumption, i.e. keeping the existing design structure in place, it is useful to consider extending the set of possible responses to include those that can act against a broad class of intruders. Indeed, T cells do produce e.g. cytokines which hinder replication of all viruses (e.g. interferon-γ). Note that they do so as part of their antigen-specific activation pattern. The question is why such "broad-spectrum" responses are produced just like specific responses, after clone-wise independent, antigen-specific detection. Can this be the optimal approach? At first sight, one could suspect that any decision to produce broad-spectrum responses should be based only on signals of an equally non-specific nature.

The correct strategy is again found by applying the usual formalism (with the standard, non-spiky background model). The derivation is unchanged up to the level of the specific u_i variables, and these may then be used as the formal signals s_i for the new, additional detection task: In this task, the null-hypothesis (no intruder) is still simple (in the statistics sense), but the alternative hypothesis is not; any of the possible ligands may be present (for simplicity, one may ignore concurrent intruders). The required test quantity then generalizes to a sum over all conditional likelihood ratios, i.e. $H = \sum_i p_i(\rho_i; u_i)/p_i(0; u_i)$. As before, the small effects of u_i-discreteness and residual Poisson-noise may be neglected, replacing u_i by h_i which has the density.

$$p_i(\rho_i; h_i) = \frac{1}{\rho_i} e^{-h_i/\rho_i} * e^{-h_i} = \frac{1}{\rho_i - 1}(e^{-h_i/\rho_i} - e^{-h_i}) \tag{24}$$

with clone-specific SNR parameters ρ_i. This yields the test quantity (up to monotonicity) as

$$H = \sum_i H_i = \sum_i \frac{\exp[h_i(\rho_i - 1)/\rho_i] - 1}{\rho_i - 1}. \tag{25}$$

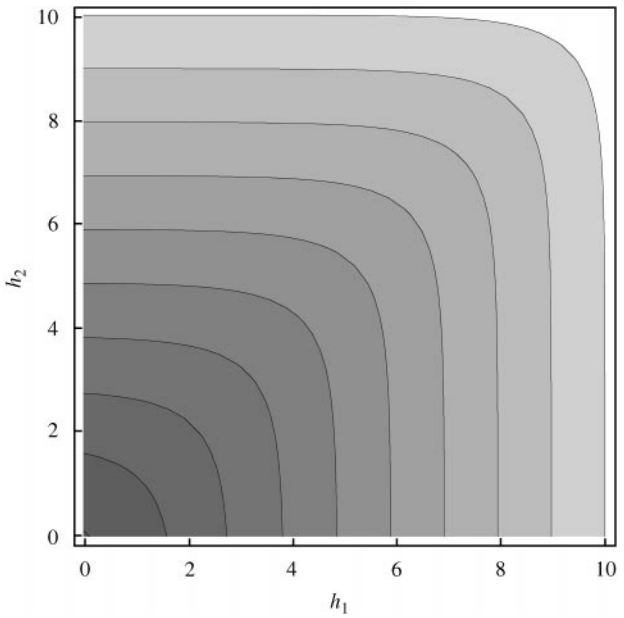

FIG. 10. Contours of the cytokine-production test quantity h as given by eqn (25), for two signals h_1, h_2, for equal $\rho_i \gg 1$. Note the almost square contour shape for reasonably large values of H. A true square-shaped detection boundary is equivalent to taking the logical OR-function of independent clone-specific detections.

The dominant terms are the ones with $\rho_i \gg 1$, i.e. those which carry the signals well above the noise. Note also that the H threshold $\theta \gg 1$, as required for an acceptably low P_f. For such large H, one can neglect all except the exponential terms in the result. The essential point to note then is that in this regime, the contour surface $H = \theta$ is almost hypercubical, except for small "rounded off" patches near the vertices. In fact, the two-dimensional contour plot, Fig. 10, illustrates that the hypercubical shape is already a good approximation for contours at θ levels well below what is needed for acceptably small P_f.

Thus, the optimal decision criterion $H > \theta$ can be replaced to good accuracy by *independent* decisions $H_i > \theta$, corresponding to the use of a hypercubical decision boundary. Since each dimension corresponds to a previously derived test quantity implemented by a specific T cell clone, the conclusion is that each cell can also independently trigger production of non-specific responses, just as it does for its specific response against the epitope which forms its cognate ligand. Note that each type of response may use its own threshold value θ, since the consequences of a false alarm may well differ between them. However, much of the machinery for computing the test quantity can be shared between the two tasks.

The resulting scheme, which fits the behaviour of real T cells, is very different from a decision based on "broad-spectrum" sensing of ligands. This latter alternative, having a decision boundary $\sum_i H_i = \theta$, would appear in Fig. 10 as contours running perpendicular to the diagonal, instead of to the axes.

5. Discussion

Applying detection theory to the task of immune surveillance, I have derived a substantial list of functional design properties required for achieving near-optimal detection performance. This "top-down design" approach consists essentially of decoding the formalism of Neyman–Pearson testing into a set of functional demands on the machinery which should compute a suitable test quantity and threshold. In contrast to other top-down modelling approaches which are often verbal, this approach specifies the required functional units mathematically. This leaves only those ambiguities which are, by construction, functionally irrelevant, and it allows one to study how the results depend quantitatively on the assumptions about the task or the available signals. It also guarantees that the many parts of the design fit together in a functionally coherent way—an aspect which easily escapes the reach of bottom-up models or experiments.

The design results are logically stratified by their dependence on a gradually introduced set of assumptions concerning the basic task and the available signals. In most cases, each new section of the paper adds or modifies one such assumption, and then derives the new functional design feature(s) which the general detection formalism dictates in the case at hand. A full summary of this sequence of constructions is impractical, but it may be helpful to mention again some of the more basic assumptions and the functional features they imply in an optimal design.

A fundamental assumption is that many distinct, unpredictable signals exist, each of which

has a noisy background level and, possibly, a contribution from an unknown intruder. The optimal detection formalism then dictates the use of a large, diverse set of detector units (cells), each of which must apply a specific filter (receptor) to the pool of available signals (ligands), and compare the result to a threshold. It also shows the need for adaptation to the cell-specific background level. Adding that the formal signals are time-series of discrete events occurring at a low rate (ligand binding events) dictates very early signal amplification, plus temporal integration. The biochemically simplest selective signal amplifier turns out to be "serial receptor triggering", which implements integration in the same go. It also allows gain-control, as required for adapting cells to their specific background. All other design features follow similarly, dictated by a few added or modified assumptions about the statistics of the available signals. This contrasts with *ad hoc* constructed models, in which the necessity or irrelevance of chosen model ingredients is unclear until the model has been fully analysed.

In this paper, I have focussed on (but not exhausted) the broad range of properties which can be derived within the general framework of optimal detector design. The functional specification for each feature has been worked out in sufficient detail for making useful comparisons of the design to reality (see below). In addition, I have explicitly computed the detection performance for some simple working versions of the design. Comparing variants with different levels of control over contact-time fluctuations has shown that this affects the net sensitivity (detection limit) of the system by up to an order of magnitude. Computing the performance of the more extended versions of the design is a viable project, but it would have exceeded the size and scope of this paper. It seems particularly worthwhile to calculate how much performance improvement is offered by co-stimulation (Section 4.1), or by the (ant)agonist type of amplification, adaptation and clonal selection for which the basic role has been established here (Section 4.2). In principle, such calculations make the model into a precise "gold-standard" for gauging the performance of real immune systems. It will be experimentally challenging to measure the required data, i.e. the probability densities of ligand levels, in particular in the background condition, but the problem is not fundamentally different from that in vision research, where a comparison to detection theory has become a standard procedure.

Since each individual design feature has already been discussed in its own section of the paper, I devote the remaining discussion to questions which are raised, or put into sharper focus, by using the design as a well-defined, independent "reference" for comparison with real immune system components. Indeed, the main virtue of deriving the design with an intentional neglect of evolutionary considerations is that it keeps the results free of an inherent bias towards merely "reproducing" real T cells from explicitly realistic assumptions. This design approach avoids the risk of a logical circularity in comparing the model to reality, and it makes finding matches or mismatches equally informative.

Thus, it is non-trivial that each of the design features matches with properties and mechanisms found in real T cells. The merely convenient labelling of the designed detectors as "T cells" clearly begged the question of whether any similarity actually exists. The most basic features (e.g. cellularity, specificity, diversity) are very well known experimentally, but deriving them from simple statistics quantifies their fundamental role in solving a highly parallel detection problem. Similar conclusions apply for many of the other features I derived, but some features seem to be less well characterized experimentally than befits their importance as identified in the design. For example, the substantial dependence of net sensitivity on the relative size of fluctuations in the contact time τ suggests that some relevant control mechanisms should exist. In other cases, e.g. high-zone tolerance, the phenomenon as such is well known, but the theory provides a clear functional reason for its existence, independent of the underlying mechanism(s).

In searching for informative matches or mismatches, one should look only at features related to detection *per se*. This excludes several components of real immune systems, even if they play an essential role in the overall success or failure of the system as a whole. For example, I have ignored the task of "preprocessing" the pathogens, i.e. of generating the ligand signals which I simply

assumed to be available to T cells. This voids any comparisons of the design to cells specialized in antigen presentation. Another case in point is the distinction between "naive" and 'memory" subtypes among the lymphocytes. By assuming here that signals are unpredictable, which also excludes any temporal correlation, I have robbed the design of any notion of time (on scales beyond τ). Thus, the design is simply mute on anything like memory cells. Clearly, this invites future generalisations of the design, e.g. by assuming the random arrival of intruders to have long-time correlations.

Comparing the design to B cells yields an interestingly mixed score: the most basic features (e.g. cellularity, specificity, diversity) match well, but the receptors of most B cells do not exploit the kinetics of "serial triggering" used by TCRs (Lanzavecchia *et al*., 1999). Thus, I expect that B cells cannot attain the ultimate sensitivity available to T cells, and that they should come into action later in the course of an infection than the T cells do. Indeed, most B cells are dependent on activated helper T cells. There is another class of B cells (B-1), which does appear to fit the design, when given a properly modified assumption: several pathogens present spatially repetitive signals on their surface. The formalism then demands that signal integration occurs in space, instead of time. This replaces "serial triggering" by its spatial analogue, as proposed informally before (Lanzavecchia *et al*., 1999).

Finding clear mismatches is both expected and informative, because they should indicate those features of real immune systems which are dominated by the evolutionary or structural constraints which I have intentionally kept out of the design. The search for mismatches leads to opposite results depending on which of two very different types of immune system one considers.

In "adaptive" immune systems (Janeway & Travers, 1996), which use a high diversity of clones, at least the T cells appear to lack clear functional mismatches with the design. Note that striking mismatches should occur if real T cells reflected an evolutionary history dominated by biochemical constraints, accidentally fixed arbitrary choices, or adaptations which only counteract a few specific subversion strategies by pathogens. However likely the occurrence of such historical events, they seem not to have left large scars as far as the basic detection function is concerned. In fact, co-evolution could well have driven immune systems towards the present near-optimal detection strategy: many immune evasion schemes used by viruses (Ploegh, 1998) subvert or inhibit the degradation of pathogens and presentation of their antigens. Reduced antigen presentation only sharpens the need for higher sensitivity of the detection stage, which may accelerate evolution towards functional equivalence with the present design.

Fundamental mismatches are found when comparing the design to the simple immune systems which occur e.g. in invertebrates. The most striking distinction is the lack of a large receptor diversity, even if the available number of cells would allow it to be larger. Only a few dozen signals seem to be specifically detected, each of which can betray the presence of a broad class of pathogens. This mismatch with the design sharpens the notion that the evolution of adaptive, high-diversity immune systems constituted a fundamental change of strategy in the defence against pathogens. Adaptive immunity appears to have arisen (Agrawal *et al*., 1998) quickly after a transposon insertion enabled recombinase activity, which generates the huge diversity of immune receptors in the T and B cells of adaptive immune systems. Without this random generator of diversity, the detection strategy I derived here would be essentially impossible. Broadly speaking, the more ancient strategy (still used by e.g. invertebrates) is to have fixed (germ-line encoded) receptors which allow detection of just a small set of signals. For this approach to stand any chance of long-term success, the only usable signals are those that are both clearly distinct from all strong background signals, and conserved under pathogen evolution. Note that this approach is quite dependent on constraints which prevent pathogens from evolving all their signals such that they either mimic the host's background, or escape from the small set of specialized receptors in the host immune system.

The very long time-scale on which non-adaptive immune systems have persisted suggests that such constraints on pathogen evolution do in fact exist. Indeed, extant versions of non-adaptive

immunity still show significant similarity to much older systems which persist in plants (Baker *et al*., 1997). Moreover, even higher animals maintain a non-adaptive immune system, but in a special form. In addition to the "adaptive" immune system, one finds an "innate" system (Hoffmann *et al*., 1999), which supplies two types of functions: one consists of a broadly reacting type of immunity analogous to its ancestral system, while the other consists of providing co-stimulation to the adaptive high-diversity system (Fearon & Locksley, 1996), signalling an infection by any of a broad class of pathogens.

The present design is in fact mute about non-adaptive immune systems, because it lacks any notion of long timescales. Still, it does reach as far as possible, in that it specifies how to use non-specific signals (Section 4.1). This defines the "interface" between the adaptive and the innate parts of the vertebrate immune system. Deriving the optimal machinery for generating these non-specific signals will require a design based on different assumptions. It may be sufficient to let the signal statistics include correlations over even longer time-scales than required for having memory-cells emerge within the high-diversity setting of the present design. The larger goal will be to derive and analyse the structure and performance of non-adaptive immunity, either in its independent or its subsumed functional role, and to compare it to the present results for adaptive immune systems.

I am very grateful to José Borghans, Paulien Hogeweg, Can Keşmir, Rob de Boer, Lee Segel, and Jorge Carneiro for providing important information, criticism, and encouragement. This work was sponsored by NWO/SLW and the Dutch AIDS Foundation (grant PccO-1317).

REFERENCES

Abramowitz, M. & Stegun, I. A. (1972). *Handbook of Mathematical Functions*. New York: Dover.

Agrawal, A., Eastman, Q. M. & Schatz, D. G. (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* **394,** 744–751.

Alam, S. M., Travers, P. J., Wung, J. L., Nasholds, W., Redpath, S., Jameson, S. C. & Gascoigne, N. R. J. (1996). T cell-receptor affinity and thymocyte positive selection. *Nature* **381,** 616–620.

Baker, B., Zambryski, P., Staskawicz, B. & Dinesh-Kumar, S. P. (1997). Signalling in plant–microbe interactions. *Science* **276,** 726–733.

Barton, G. M. & Rudensky, A. Y. (1999). Requirement for diverse, low-abundance peptides in positive selection of T cells. *Science* **283,** 67–70.

Borghans, J. A. M. & deBoer, R. J. (1998). Crossreactivity of the T-cell receptor. *Immunol. Today* **19,** 428–429.

Borghans, J. A. M., Noest, A. J. & deBoer, R. J. (1999). How specific should immunological memory be? *J. Immunol.* **163,** 569–575.

deBoer, R. J. & Perelson, A. S. (1993). How diverse should the immune system be? *Proc. Roy. Soc. Lond. B* **252,** 171–175.

Fearon, D. T. & Locksley, R. M. (1996). The instructive role of innate immunity in the acquired immune response. *Science* **272,** 50–53.

Grossman, Z. & Paul, W. E. (1992). Adaptive cellular interactions in the immune system: the tunable activation threshold and the significance of subthreshold responses. *Proc. Natl Acad. Sci. U.S.A.* **89,** 10 365–10 369.

Hoffmann, J. A., Kafatos, F. C., Janeway Jr., C. A. & Ezekowitz, R. A. B. (1999). Phylogenetic perspectives in innate immunity. *Science* **284,** 1313–1318.

Janeway, C. A. & Travers, P. (1996). *Immunobiology*. London: Current Biology Ltd.

Langman, R. E. & Cohn, M. (1996). A short history of time and space in immune discrimination. *Scand. J. Immunol.* **44,** 544–548.

Lanzavecchia, A., Iezzi, G. & Viola, A. (1999). From TCR engagement to T cell activation: a kinetic view of T cell behaviour. *Cell* **96,** 1–4.

Margulies, D. H. (1996). An affinity for learning. *Nature* **381,** 558–559.

Matis, L. A., Glimcher, L. H., Paul, W. E. & Schwartz, R. H. (1983). Magnitude of response of histocompatibility-restricted T-cell clones is a function of the product of the concentrations of antigen and Ia molecules. *Proc. Natl Acad. Sci. U.S.A.* **80,** 6019–6023.

McKeithan, T. W. (1995). Kinetic proofreading in T-cell receptor signal transduction. *Proc. Natl Acad. Sci. U.S.A.* **92,** 5042–5046.

Mitchison, N. A. (1964). Induction of immunological paralysis in two zones of dosage. *Proc. Roy. Soc. Lond.* **161,** 275–292.

Neumeister Kersh, E., Shaw, A. S. & Allen, P. M. (1998). Fidelity of T cell activation through multistep T cell receptor ζ phosphorylation. *Science* **281,** 572–575.

Ploegh, H. L. (1998). Viral strategies of immune evasion. *Science* **280,** 248–253.

Rabinowitz, J. D., Beeson, C., Lyons, D. S., Davis, M. M. & McConnell, H. M. (1996). Kinetic discrimination in T-cell activation. *Proc. Natl Acad. Sci. U.S.A.* **93,** 1401–1405.

Rothenberg, E. V. (1996). How T cells count. *Science* **273,** 78–79.

Segel, L. A. & Perelson, A. S. (1988). Computations in shape-space: a new approach to immune network theory. In: *Theoretical Immunology II*; *SFI Studies in the Science of Complexity*, Vol. III. pp. 321–343. Redwood City, CA: Addison-Wesley.

SHANNON, C. (1948). The mathematical theory of communication. *Bell System Tech. J*. **27,** 379–423 and 623–656.

SLOAN-LANCASTER, J. & ALLAN, P. M. (1996). Altered peptide ligand-induced partial T cell activation: molecular mechanisms and role in T cell biology. *Ann. Rev. Immunol*. **14,** 1–27.

VALITUTTI, S., MULLER, S., CELLA, M., PADOVAN, E. & LANZAVECCHIA, A. (1995). Serial triggering of many T-cell receptors by a few peptide-MHC complexes. *Nature* **375,** 148–151.

VIOLA, A. & LANZAVECCHIA, A. (1996). T cell activation determined by T cell receptor number and tunable thresholds. *Science* **273,** 104–106.

VIOLA, A., SCHROEDER, S., SAKAKIBARA, Y. & LANZAVECCHIA, A. (1999). T lymphocyte costimulation mediated by reorganisation of membrane microdomains. *Science* **283,** 680–682.

WHALEN, A. D. (1971). *Detection of Signals in Noise*. New York: Academic Press.

WÜLFING, C. & DAVIS, M. M. (1998). A receptor/cytoskeletal movement triggered by costimulation during T cell activation. *Science* **282,** 2266–2269.