S1 Appendix: Supporting Information for the paper
“Multiplexing information flow through dynamic
signalling systems”

Giorgos Minas, Dan Woodcock, Louise Ashall, Claire Harper, Mike White, David Rand

July 20, 2020

Contents

1 Mathematical details and notation 3

2 Fisher Information, Sensitivity Matrix & Multiplexing Capacity 4
   2.1 Fisher Information Matrix & KL Divergence 4
   2.2 Information geometry 5
   2.3 The multiplexing sensitivity matrix $s$ 5
   2.4 Multiplexing Capacity 5
      2.4.1 Finding optimal multiplexing sets 6
   2.5 Sensitivity analysis for multivariate normal distributions 7
      2.5.1 The matrix $s$ characterises sensitivity 8
      2.5.2 Optimality of $s$ 9
      2.5.3 The multiplexing capacities of a model 9

3 Reaction Networks & Stochastic NF-$\kappa$B model 10
   3.1 System size parameter 10
   3.2 The classical deterministic approximation 11
   3.3 Stochastic NF-$\kappa$B model 11
   3.4 The mNF-$\kappa$B model 15
   3.5 Another model for NF-$\kappa$B modification 17

4 LNA, pclLNA and calculating the FIM 21
   4.1 The linear noise approximation (LNA) 21
   4.2 Phase corrected LNA (pclLNA) 22
      4.2.1 The pclLNA distribution 23
      4.2.2 The pclLNA joint distribution of multiple transversals 24
   4.3 pclLNA simulation 26
   4.4 Channel capacity estimation 27
   4.5 Channel capacity under the stochastic model of NF-$\kappa$B in Tay et al. [16] 27
   4.6 Simulation of EGR1 and COX-2 genes 28
   4.7 Fisher Information Matrix (FIM) under the pclLNA 29
      4.7.1 Derivatives for FIM computation 29
   4.8 FIM computation for pclLNA joint distributions 31
   4.9 Effects of observing less time-points on the singular values of the FIM 31

5 NF-$\kappa$B target genes 32
List of Figures

A  Base NF-κB model: ODE system solutions ........................................... 15
B  mNF-κB model: ODE system solutions .................................................. 17
C  amNF-κB model: Diagram .................................................................... 18
D  amNF-κB model: ODE system solutions .................................................. 20
E  The amNF-κB model analysis. ................................................................. 21
F  The main steps in the pcLNA simulation algorithm. ................................. 27
G  Histograms of data simulated by the algorithm in [16] ............................. 28
H  Singular values for the less time-points .................................................. 32
I  Microarray data 1 ................................................................................. 33
J  Microarray data 2 ................................................................................. 34
K  Gene expression summary table ............................................................ 35

List of Tables

A  The backward reduction QR decomposition used in section 2.4.1 .............. 7
B  Base NF-κB model: variables, ODE system initial conditions .................. 12
C  Base NF-κB model: ODE system ............................................................ 12
D  Reactions of the base NF-κB system and their rates. ............................... 13
E  The parameters of the base NF-κB system and the values used to derive their
   ODE solution......................................................................................... 14
F  mNF-κB model: ODE system variables .................................................... 15
G  mNF-κB model: ODE system ................................................................. 16
H  mNF-κB model: extra parameters ........................................................... 16
I  amNF-κB model: variables and initial conditions .................................... 18
J  amNF-κB model: ODE system ................................................................. 19
K  amNF-κB model: extra parameters ........................................................ 20
In the following, we provide technical details and various supplementary illustrations to those provided in the main paper that hereafter is referred as I.

1 Mathematical details and notation

**Transpose of a vector or matrix and inner product** We denote $a^T$ and $A^T$ the transpose of the vector $a \in \mathbb{R}^n$ and matrix $A \in \mathbb{R}^{n \times n}$, respectively. We use the convention to write vectors as columns, i.e. $a \in \mathbb{R}^{n \times 1}$, and therefore the transpose, $a^T$, is a row vector, i.e. $a^T \in \mathbb{R}^{1 \times n}$. The inner or dot product of two vectors $a, b \in \mathbb{R}^n$ is $a^T b = a_1 b_1 + \cdots + a_n b_n$.

**Vectorisation of a matrix and the Kronecker product** The vectorisation operator $\text{vec}(\cdot)$ applied to an $m \times n$ matrix $A = (a_{ij})$, maps $A$ to an $mn \times 1$ vector, $\text{vec}(A) = (a_{11}, \ldots, a_{1n}, \ldots, a_{m1}, \ldots, a_{mn})^T$.

The Kronecker product of an $m \times n$ matrix $A = (a_{ij})$ and an $s \times t$ matrix $B$ is the $ms \times nt$ matrix $A \otimes B = (a_{ij}B)_{ij}$. Then we have the following very useful equality: $\text{vec}(ABC) = (C^T \otimes A)\text{vec}(B)$ (see [7]).

**Probability density functions** For a stochastic process $\{X(t), t \geq 0\}$ with continuous state space, we write the probability density function (pdf) at time $t \geq 0$ as $P(X(t))$. The corresponding conditional pdf of $X(t)$, $t > t_0$, given a fixed initial state $X(t_0) = x_0$ or an initial distribution $X(t_0) \sim P_0$ are respectively denoted as $P(X(t)|X(t_0) = x_0)$ and $P(X(t)|X(t_0) \sim P_0)$. The indicator LNA, as in $P_{\text{LNA}}$, is used in probability functions such as the above to denote that these are derived under the linear noise approximation (LNA).

**Multivariate Normal distribution** We write $\zeta \sim \text{MVN}(m, V)$ or $P$ is $\text{MVN}(m, V)$ to mean that the random variable $\zeta$ or the probability distribution $P$, respectively, is multivariate normal with mean vector $m$ and covariance matrix $V$.

**Singular Value Decomposition (SVD)** Singular Value Decomposition (SVD) gives a decomposition of the matrix $A \in \mathcal{F}_{m \times k}$ into a product of the form $A = UDV^T$ where $U$ is a $m \times k$ column-orthogonal matrix ($U^T U = I_k$), $V$ is a $k \times k$ orthogonal matrix ($V^T V = VV^T = I_k$) and $D = \text{diag}(\sigma_1, \ldots, \sigma_k)$ is a diagonal matrix. This is the version of SVD that is often called thin SVD (see [14] for more details). The SVD applies for matrices defined in arbitrary fields $\mathcal{F}$, but henceforth we only consider real matrices, i.e. $\mathcal{F} = \mathbb{R}$. The ordered elements $\sigma_1 \geq \cdots \geq \sigma_k$ are the singular values of $A$. We write the $i$th singular value $\sigma_i = \sigma_i(A)$. We note that the columns $V_j$, $j = 1, \ldots, k$ of $V$ form an orthonormal basis for $\mathbb{R}^k$. If the last $n$ of the singular values are zero then the last $n$ columns are orthonormal basis for the kernel of $A$. If $m < k$ then $n \geq k - m$.

In what follows we will refer to the column vectors $V_j = (V_{ij}, \ldots, V_{kj})^T$ and $U_j = (U_{i1}, \ldots, U_{mj})^T$, $j = 1, 2, \ldots, k$ as the right and left singular vectors respectively, for which $AV_j = \sigma_j U_j$ and $U_j^T A = \sigma_j V_j^T$, with $V_j^T$ the $j$th row of $V$.

**QR-decomposition (QR)** The QR-decomposition with column pivoting produces a decomposition of any $m \times k$ matrix $M$ as $MP = QR$ where $P$ is a permutation matrix, which is unitary ($P^T P = PP^T = I$), $Q$ is an $k \times k$ orthogonal matrix with orthonormal columns $Q_j$, $j = 1, \ldots, k$, i.e. $Q_i^T Q_i = 0$, if $j \neq i$ and $Q_i^T Q_i = 1$, and $R$ is an upper-triangular $m \times k$ matrix. If the row rank of $M$ is $q$ then only the first $q$ of these entries are positive.

The permutation matrix $P$ simply reorders the columns $M_i$ of $M$ so that the matrix $MP = [M_{i_1} \cdots M_{i_k}]$ with the index set $\{i_1, \ldots, i_k\} = \{1, \ldots, k\}$. It is easy to check that $M_{ij} = QR_{ij}$ where $R_{ij}$ the $j$th column of $R$. The entries $R_{ij}$ of $R$ are $R_{ij} = Q_i^T M_{ij}$, when
We next outline an approach to sensitivity analysis and multiplexing. The key point will be
\[
d \quad \text{where} \quad D
\]
That is, the KL divergence \( p \) with respect to the Fisher Information Matrix (FIM), \( \mathcal{I}(\mathbf{S}) \), of the probability distribution \( P_\mathbf{S} \) at \( \mathbf{S} \in \mathcal{S} \) is a symmetric positive-(semi)definite \( s \times s \) matrix with entries
\[
\mathcal{I}_{ij} = \mathbb{E}_{P_\mathbf{S}} (\partial_i \ell \cdot \partial_j \ell) = -\mathbb{E}_{P_\mathbf{S}} (\partial^2_{ij} \ell)
\]
Here \( \mathbb{E}_{P_\mathbf{S}} \) denotes the expectation function under the distribution \( P_\mathbf{S} \), \( \partial_i \) the partial derivative with respect to the \( i \)th component \( S_i \) evaluated at \( \mathbf{S} \in \mathcal{S} \) and \( \partial^2_{ij} \) the corresponding second order derivative.

If the signals \( S_i, i = 1, \ldots, s \), act through changing parameters \( \theta_j \) of the system (i.e. so that \( \theta = \theta(\mathbf{S}) \)) and \( \mathcal{I}_\theta \) the FIM with respect to \( \theta = (\theta_1, \ldots, \theta_n)' \) we have that
\[
\mathcal{I}(\mathbf{S}) = d\theta(S)^T \mathcal{I}_\theta d\theta(S)
\]
where \( d\theta(S) \) is the derivative of \( \theta \) with respect to \( \mathbf{S} \) at \( \mathbf{S} \in \mathcal{S} \). This well-known property of the FIM follows from the chain rule for derivatives.

The Kullback–Leibler (KL) divergence of two probability distributions \( P \) and \( Q \) with density functions \( p(x) \) and \( q(x) \), \( x \in \mathcal{X} \), is given by
\[
D_{\text{KL}}(P \| Q) = \int_{\mathcal{X}} p(x) \log \frac{p(x)}{q(x)} \, dx.
\]
That is, the KL divergence \( D_{\text{KL}}(P \| Q) \) is the expected value of the likelihood ratio \( \lambda(x) = \log p(x)/q(x) \) with the expectation taken with respect to \( P \) with the usual conventions when \( p(x) = 0 \) or \( q(x) = 0 \). If \( P = P_{\mathbf{S}+\delta \mathbf{S}} \) and \( Q = P_{\mathbf{S}} \), then (see [3]),
\[
D_{\text{KL}}(P_{\mathbf{S}+\delta \mathbf{S}} \| P_{\mathbf{S}}) = \frac{1}{2} \delta \mathbf{S}^T \mathcal{I}(\mathbf{S}) \delta \mathbf{S} + O(\|\delta \mathbf{S}\|^3).
\]
That is, the FIM is the hessian matrix of the above KL divergence at \( \mathbf{S} \in \mathcal{S} \).
2.2 Information geometry

If the Fisher information matrix $\mathcal{I}(S_0)$, $S_0 \in S$, is positive definite, it can also be used to define a Riemannian metric over the space of probability distributions. Firstly, we use it to define a metric on the signal space $S$ at $S_0$ by

$$\langle \delta S, \delta S' \rangle_{S_0} = \sum_{i,j} \delta S_i \delta S'_j \mathcal{I}_{ij}(S_0) = \delta S^T \mathcal{I}(S_0) \delta S'.$$

This inner product is interesting to us because,

$$D_{KL}(P_{S_0 + \delta S} || P_{S_0}) = \frac{1}{2} \langle \delta S, \delta S \rangle_{S_0} + O(||\delta S||^3).$$

Therefore, assuming that $\mathcal{I}(S)$ depends in a $C^1$ fashion on $S$, if $S$ and $S'$ are perturbations of $S_0$

$$D_{KL}(P_S || P_{S'}) = \frac{1}{2} \langle S - S', S - S' \rangle_{S_0} + O(\max\{\|S - S'\|^3, \|S - S_0\|^3, \|S' - S_0\|^3\}). \quad (2)$$

If this is the case we say $D_{KL}(P_S || P_{S'}) = \frac{1}{2} \langle S - S', S - S' \rangle_S$ up to third order.

2.3 The multiplexing sensitivity matrix $s$

Since the FIM $\mathcal{I} = \mathcal{I}(S_0)$ is symmetric and positive semi-definite its SVD is of the form $VDV^T$ where $V$ is orthogonal and $D$ is diagonal with entries $\sigma_1^2 \geq \cdots \geq \sigma_s^2 \geq 0$. The sensitivity matrix $s$ associated with $\mathcal{I}$ is $s = D^{1/2} V$. We explain the reason for this name below.

2.4 Multiplexing Capacity

Recall the definition of $D_{KL}^{(i,S_0)}$ as $\min_s \eta_i^{-2} \min_{S'} D_{KL}(P_S || P_{S'})$ where the $i$th component of $S - S_0$ has magnitude $\eta_i = \|S_i - S_0,i\| > 0$ and the minimum is taken over all perturbations $S'$ of $S_0$ where the $i$th component $S_i$ has not changed, i.e. $S'_i - S_{0,i} = 0$. Note that $D_{KL}^{(i,S_0)}$ depends upon the set of signals $(S_1, \ldots, S_k)$, $k \leq s$, being considered. If we need to stress this we will write $D_{KL}^{(i,S_0)}(S_1, \ldots, S_k)$ instead. Note that in general $D_{KL}^{(i,S_0)}(S_1, \ldots, S_k) \neq D_{KL}^{(i,S_0)}(S_1, \ldots, S_{k-1})$.

Given any set of integers $1 \leq i_1 < \cdots < i_r < s$ distinct from $i$ consider the columns $s_{i_j}$ of $s$. Denote by $n(s_i|s_{i_1}, \ldots, s_{i_r}) = n(i|i_1, \ldots, i_r)$ the unique vector $n$ normal to the vectors $s_{i_1}, \ldots, s_{i_r}$ such that $s_i - n$ lies in the span of $s_{i_1}, \ldots, s_{i_r}$. That is,

$$n^T s_j = 0, \quad j = 1, \ldots, r, \quad \text{and} \quad n^T (s_i - n) = 0. \quad (3)$$

In the case where $1 \leq i_1 < \cdots < i_r < s$ consists of all $j$ with $j \neq i$ we use the notation $n(s_i|s_{j}, j \neq i)$. We now prove that, up to third order terms,

$$D_{KL}^{(i,S_0)} = \frac{1}{2} \|n(s_i|s_{j}, j \neq i)\|^2. \quad (4)$$

We first prove (4) for the case where $i = s$. Let $s = QR$ be the $QR$ decomposition of $s$. By definition $n = n(s_i|s_j, j \neq s) = R_{ss} Q_s$, and therefore $\|n\| = \|R_{ss}\|$. To check this, one can write $n = \sum c_j Q_j$ and use (3) to derive the coefficients $c_j = 0$, $j < s$ and $c_s = R_{ss}$.

Then, see that $\mathcal{I}(S_0) = s^T s = R^T R = V D V^T$ since $Q^T Q$ is the identity matrix. But, by (2), up to third order,

$$2D_{KL}(P_S || P_{S'}) = \langle \delta S, \delta S \rangle_{S_0} = \delta S^T \mathcal{I}(S_0) \delta S = \delta S^T R^T R \delta S = \|R \delta S\|^2. \quad (5)$$
where $\delta S = S - S'$. Because $R$ is upper triangular and $\delta S_s = S_s$ the last entry of the vector $R\delta S$ is $R_{ss}S_s$. It follows that, up to third order terms,

$$2D_{KL}(P_S||P_{S'}) = R_{ss}^2S_s^2 + \sum_{j=1}^{s-1} (R\delta S)_j^2 \geq |R_{ss}S_s|^2 = \|n\|^2\eta_s^2,$$

where $(R\delta S)_j$, the $j$th entry of the vector $R\delta S$ and the minimum is achieved when $S'_j$ is such that $(R\delta S)_j = 0$, for $j = 1, \ldots, s - 1$. These $S'_j$’s can be found if we let $R'$ be the $(s - 1) \times (s - 1)$ upper triangular matrix obtained by deleting the last row and column from $R$ and let $R_k$ be the $(s - 1)$-dimensional vector obtained by deleting the last element from the last column of $R$. Then, if $S'$ is the $(s - 1)$-dimensional vector solving $R'\delta S' = -R_k$ and $\delta S$ is the $s$-dimensional vector obtained by adding $\delta S_s$ to the end of $\delta S'$ we have that

$$\|R\delta S\|^2 = |\delta S_s|^2.$$  

Therefore, up to third order terms, $2D_{KL}(i,S_0) = R_{ss}^2$.

In case $i < s$, we can simply use a permutation matrix, $P$, such that the last column of the matrix $sP$ is $s_s$, and follow the same steps as above to show that, up to third order terms, $2D_{KL}(i,S_0) = |n(s_i|s_j, j \neq i)|^2 = R_{ss}^2$, where $R_{ss}$ the bottom right entry of the $R$ matrix that satisfies $sP = QR$. This proves $4$.  

### 2.4.1 Finding optimal multiplexing sets

In view of the above result we ask the following question, given a set of signals $S_1, \ldots, S_s$, for which subsets $S_i_1, \ldots, S_i_r$, $r \leq s$, can we effectively multiplex in the sense that all the multiplexing capacities for $S_i_1, \ldots, S_i_r$ are above a given threshold.

We first choose the column pivoting matrix $P^*$ so that the diagonal entries, $R_{jj}$, of the $R$ matrix in the QR decomposition $sP^* = [s_1; \ldots; s_s] = QR$ give the minimum multiplexing capacities $MX(S_i_1; \ldots; S_i_r)$. That is, up to third order,

$$MX(S_i_1, \ldots, S_i_k) = \min_{i = i_1, \ldots, i_k} D_{KL}(i,S_0)(S_i_1, \ldots, S_i_k) = R_{kk}^2/2, \quad k = 1, \ldots, s. \quad (6)$$

This QR decomposition can be easily derived by an algorithm such as this provided in Table A. The idea is to start with the full set of signals and use QR to find and remove from the current set of signals, say $S_k = \{S_{i_1}, \ldots, S_{i_k}\}$, $k = s, s - 1, \ldots, 2$, the signal that gives the minimum multiplexing capacity $\min_{i=i_1,\ldots,i_k} \|n(s_i|s_j, j \in S_k \setminus s_i)\|^2$. These multiplexing capacities can be derived by the following steps: (i) pivoting through an appropriate $P$ the $s_j$ columns, $l = 1, \ldots, k$, to get $sP$ with last column equal to $s_{i_l}$, (ii) performing QR decomposition to get $sP = QR$ and (iii) using that $R_{kk}^2 = \|n(s_i|s_l, S_{i_l} \setminus s_i)\|^2$ to compute the minimum $D_{KL}$ (see Table A). Note that this is a backward reduction QR decomposition as opposed to the standard (forward addition) column pivoting QR decomposition provided by software such as MATLAB [11].

It is easy to see that, as in the standard column pivoting QR decomposition, the diagonal entries $R_{ii}$ of $R$ in $sP^* = QR$ are in decreasing order

$$R_{11}^2 \geq \cdots \geq R_{ss}^2.$$ 

This is because, for $l \leq s$,

$$|R_{ll}| = \|n(s_l|s_1, \ldots, s_l-1)\| \leq \|n(s_{i_{l-1}}|s_1, \ldots, s_{i_{l-2}}, s_{i_l})\| \leq \|n(s_{i_{l-1}}|s_{i_1}, \ldots, s_{i_{l-2}})\| = |R_{i_{l-1},i_{l-1}}|.$$ 

However, the backward reduction QR, unlike the forward addition QR, ensures that the diagonal entries of $R$ can be used to derive the minimum multiplexing capacities as in $6$.

Once the appropriate ordering of signals $S_i_1, \ldots, S_i_s$ is identified by the backward reduction QR, we can perform a QR decomposition without pivoting to the sensitivity matrix $s^* = [s_1; \ldots; s_s] = QR$ and by examining the diagonal entries of $R$, we can see
which signals $S_{i_1}, S_{i_2}, \ldots, S_{i_k}$, $k = 1, \ldots, s$ can multiplex before the multiplexing capacity
$\text{MX}(S_{i_1}, S_{i_2}, \ldots, S_{i_k}) = R^2_{kk}/2$ falls below a threshold. Equivalently, one can start from the
full set of signals $S_1, \ldots, S_s$ and find the minimum subset of signals, $S_{i_{s-r+1}}, \ldots, S_{i_s}$, $r \geq 0$,
that needs to be removed (i.e. keep unperturbed) to derive a multiplexing subset, i.e. a subset
that has $\text{MX}(S_{i_1}, S_{i_2}, \ldots, S_{i_{s-r}}) = R^2_{s-r,s-r}/2$ larger than a threshold.

1. Compute the SVD, $\mathcal{I} = V D V^T$, of the FIM $\mathcal{I}$ and set $\tilde{s} = D^{1/2} V^T$ and $s' = s$
2. For $i = 1, \ldots, s'$:
   (a) Use a column pivoting matrix $P$ to get $\tilde{s}' = \tilde{s} P$ so that its last column
       $\tilde{s}'_s = \tilde{s}_s$, where $\tilde{s}_i$ the $i$th column of $\tilde{s}$.
   (b) Perform QR decomposition to compute $QR = \tilde{s}'$
   (c) Set $r_i = |R_{s's'}|
3. Find $i_{s'} = \arg\min_{i=1,\ldots,s'} r_i$
4. Set $s' = s' - 1$ and $\tilde{s} = [\tilde{s}_1 \cdots \tilde{s}_{i_{s'}} - 1 \tilde{s}_1, \ldots, \tilde{s}_s]$
5. If $s' > 1$ return to step 2.

Table A: The backward reduction QR decomposition used in section 2.4.1

2.5 Sensitivity analysis for multivariate normal distributions
In this section we assume that $P_\mathcal{S}$ is MVN$(\mu, \Sigma)$, with mean $\mu = \mu(\mathcal{S})$, and covariance
$\Sigma = \Sigma(\mathcal{S})$. Then the entries of the FIM are

$$\mathcal{I}_{ij}(\mathbf{S}_0) = (\partial_i \mu)^T \Sigma^{-1} (\partial_j \mu) + \frac{1}{2} \text{tr}(\Sigma^{-1} (\partial_i \Sigma) \Sigma^{-1} (\partial_j \Sigma))$$

(7)

where all derivatives are taken at $\mathbf{S} = \mathbf{S}_0$. This can also be written using vec notation as

$$\mathcal{I}_{ij}(\mathbf{S}_0) = (\partial_i \mu)^T \Sigma^{-1} (\partial_j \mu) + \frac{1}{2} \text{vec}(\partial_i \Sigma)(\Sigma^{-1} \otimes I)(I \otimes \Sigma^{-1})\text{vec}(\partial_j \Sigma).$$

Now consider the $N \times s$ matrix $(N = n + n^2)$

$$\mathcal{L} = \begin{pmatrix}
\frac{\partial \mu}{\partial \text{vec}(\Sigma)} \\
\frac{\partial_1 \mu}{\partial \text{vec}(\Sigma)} & \ldots & \frac{\partial_k \mu}{\partial \text{vec}(\Sigma)}
\end{pmatrix},$$

(8)

which is the derivative (linearisation matrix) of the mapping $\mathbf{S} \mapsto \Theta(\mathbf{S}) = (\mu(\mathbf{S}), \text{vec}(\Sigma(\mathbf{S})))$
at $\mathbf{S}_0$. That is, if we let $\delta \mu = (\delta \mu_1, \ldots, \delta \mu_n)^T$ and $\delta \text{vec}(\Sigma) = (\delta \Sigma_{11}, \ldots, \delta \Sigma_{nn})^T$, with
$\delta \mu_i = \mu_i(S + \delta S) - \mu_i(S)$ and $\delta \Sigma_{ij} = \Sigma_{ij}(S + \delta S) - \Sigma_{ij}(S)$, then

$$\delta \Theta = (\delta \mu, \delta \text{vec}(\Sigma))^T = \mathcal{L} \delta \mathbf{S} + O(\|\delta \mathbf{S}\|^2).$$

(9)

If we also define the $N \times N$ matrix $\mathcal{F}$ as the Cholesky decomposition of the positive-definite matrix $F$ (i.e. $\mathcal{F}^T \mathcal{F} = F$), where

$$\mathcal{F} = \begin{pmatrix}
\Sigma^{-1} & 0 \\
0 & (\Sigma^{-1} \otimes I)(I \otimes \Sigma^{-1})/2
\end{pmatrix}$$

then we can write the Fisher information in (7) as

$$\mathcal{I}(\mathbf{S}_0) = (\mathcal{F} \mathcal{L})^T (\mathcal{F} \mathcal{L}).$$

Therefore, $\mathcal{F} \mathcal{L}$ is a linear map from $\mathcal{S}$ to $\mathbb{R}^N$ which sends the $\langle \cdot, \cdot \rangle_{\mathbf{S}_0}$ metric to the standard one in $\mathbb{R}^N$:

$$\langle \delta \mathbf{S}, \delta \mathbf{S}' \rangle_{\mathbf{S}_0} = \delta \mathbf{S}^T \mathcal{I}(\mathbf{S}_0) \delta \mathbf{S}' = (\mathcal{F} \mathcal{L} \delta \mathbf{S})^T (\mathcal{F} \mathcal{L} \delta \mathbf{S}') = \delta \mathbf{S}^T (\mathcal{F} \mathcal{L})^T (\mathcal{F} \mathcal{L}) \delta \mathbf{S}'$$

and relates the FI metric in $\mathcal{S}$ to the standard one in $\mathbb{R}^{n+n^2}$:

$$\|\delta \mathbf{S}\|^2_{\mathcal{S}} = \|\mathcal{F} \mathcal{L} \delta \mathbf{S}\|^2.$$
2.5.1 The matrix $\mathbf{s}$ characterises sensitivity

The sensitivity to changes $\delta \mathbf{S}$ in $\mathbf{S}$ of the probability distribution $\text{MVN}(\mu(\mathbf{S}), \Sigma(\mathbf{S}))$ can therefore be studied using the vector $\mathcal{F} \mathcal{L} \delta \mathbf{S}$. Equation (9) shows that

$$\mathcal{F} \delta \Theta = \mathcal{F}(\delta \mu, \delta \text{vec}(\Sigma))^T = \mathcal{F} \mathcal{L} \delta \mathbf{S} + O(\|\delta \mathbf{S}\|^2).$$

We now consider the (thin) SVD $\mathcal{F} \mathcal{L} = \mathbf{W} \mathbf{D} \mathbf{V}^T$ (as described in section 1) to deduce some useful facts about the vector $\mathcal{F} \mathcal{L} \delta \mathbf{S}$. Firstly, note that the $N \times 1$ orthogonal column vectors $\mathbf{W}_i$ of $\mathbf{W}$ and the $s \times 1$ column vectors $\mathbf{V}_i$ of $\mathbf{V}$, satisfy $\mathcal{F} \mathcal{L} \mathbf{V}_i = \sigma_i \mathbf{W}_i$, where $\sigma_i$ the singular vectors of $\mathcal{F} \mathcal{L}$. If we define $\mathbf{U}_i = \mathbf{F}^{-1} \mathbf{W}_i$, $i = 1, 2, \ldots, s$ then

$$\mathcal{L} \mathbf{V}_i = \sigma_i \mathbf{U}_i \quad \text{(10)}$$

where the $N = (n + n^2)$-dimensional vectors $\mathbf{U}_i$ can be written as $\mathbf{U}_i = (\mathbf{U}^\mu_i, \mathbf{U}^\Sigma_i)^T$ to reflect the correspondence of the first $n$ entries to the mean vector $\mu$ and the last $n^2$ entries to the covariance matrix $\Sigma$.

By (10), up to terms that are $O(\|\delta \mathbf{S}\|^2)$,

$$\delta \Theta = \mathcal{L} \cdot \delta \mathbf{S} = \mathcal{L} \cdot \sum_i (V_i^T \cdot \delta \mathbf{S}) \mathbf{V}_i = \sum_i (\sigma_i V_i^T \cdot \delta \mathbf{S}) \mathbf{U}_i = \sum_i \left( \sum_j s_{ij} \delta \mathbf{S}_j \right) \mathbf{U}_i = \mathbf{U} \cdot \mathbf{s} \cdot \delta \mathbf{S}$$

i.e.

$$\mathcal{L} = \mathbf{U} \cdot \mathbf{s} \quad \text{(11)}$$

and therefore

$$\|\delta \Theta\|_{\mathbf{s}_0} = \|\mathbf{s} \cdot \delta \mathbf{S}\| + O(\|\delta \mathbf{S}\|^2)$$

since the $\mathbf{U}_i$ are orthonormal in the $(\cdot, \cdot)_{\mathbf{s}_0}$ metric and the coefficient of $\mathbf{U}_i$ in $\mathbf{U} \cdot \mathbf{s} \cdot \delta \mathbf{S}$ is the $i$th coordinate of $\mathbf{s} \cdot \delta \mathbf{S}$.

These equations are the reason we call $\mathbf{s} = D^{1/2} \mathbf{V}^T$ the sensitivity matrix. It describes how the distribution changes as the signal $\mathbf{S}$ is changed.

Similarly, up to terms that are $O(\|\delta \mathbf{S}\|^2)$,

$$\mathcal{F}(\delta \Theta) = \sum_{i=1}^s \sum_{j=1}^s \mathbf{W}_i \mathbf{s}_{ij} \delta \mathbf{S}_j = \mathbf{W} \cdot \mathbf{s} \cdot \delta \mathbf{S} \quad \text{(12)}$$

i.e. $\mathcal{F} \mathcal{L} = \mathbf{W} \cdot \mathbf{s}$. We can now make a few useful observations:

1. Since the FIM is $\mathcal{I}(\mathbf{S}_0) = \mathbf{s}^T \mathbf{s}$, the eigenvalues of $\mathcal{I}(\mathbf{S}_0)$ are the squares of the singular values $\sigma_i$ of $\mathbf{s}$ and the right singular vectors of $\mathbf{s}$, $\mathbf{V}_i$, are the eigenvectors of $\mathcal{I}$. Further, since the column vectors $\mathbf{V}_i$ of $\mathbf{V}$ are unit vectors the squared norm of the $i$th row of $\mathbf{s}$, $\sum_j s_{ij}^2$ is $\sigma_i^2$.

2. Equation (12) shows that the change in the probability distribution $\text{MVN}(\mu(\mathbf{S}), \Sigma(\mathbf{S}))$ produced by a change $\mathbf{S} \rightarrow \mathbf{S} + \delta \mathbf{S}$, according to the Fisher information metric, is a weighted sum of the vectors $\mathcal{F}(\mathbf{U}^\mu_j, \mathbf{U}^\Sigma_j)^T$ with weights-coefficients $\sigma_j V_j^T \delta \mathbf{S}$. The change in the probability distribution is reflected to the mean through the $\mathbf{U}^\mu_j$ directions and the covariance matrix through the $\mathbf{U}^\Sigma_j$ directions.

3. The coefficients $\sigma_j V_j^T \delta \mathbf{S}$ are proportional to the singular values $\sigma_j$ and the inner products $V_j^T \delta \mathbf{S}$. The latter are the coordinates of $\delta \mathbf{S}$ in the orthonormal basis of $\mathbf{S}$ defined by the columns of the matrix $\mathbf{V}$. Therefore, if the singular values are chosen in non-increasing order, i.e. $\sigma_1 \geq \cdots \geq \sigma_s$, the largest change in the probability distribution, subject to fixed $\|\delta \mathbf{S}\|$, occurs when the change $\delta \mathbf{S}$ is parallel to $\mathbf{V}_1$ that corresponds to the largest singular value $\sigma_1$. If the singular values decay fast, there are only a few directions of the signal space that can produce a relatively large change in the $\text{MVN}(\mu(\mathbf{S}), \Sigma(\mathbf{S}))$ distribution (subject to fixed $\|\delta \mathbf{S}\|$).
4. Furthermore, the overall contribution of each coordinate $\delta S_i$ of $\delta S$ in the change of the probability distribution is measured, according to the Fisher information metric, by
\[
\sum_{j=1}^s W_jS_{ij} = W_jS_i.
\]
That is, if $\delta S = e_i$, with $e_i \in \mathbb{R}^s$ the usual unit vector with only non-zero entry $e_{ii} = 1$, then the corresponding change in the probability distribution $\text{MVN}(\mu(S), \Sigma(S))$ according to the Fisher information metric is
\[
\epsilon \mathcal{F} L e_i = \epsilon W S_i.
\]

2.5.2 Optimality of $g$

The above sensitivity matrix is optimal for capturing as much sensitivity as possible in the low order principal components of $\delta \theta = (\delta \mu, \delta \text{vec}(\Sigma))^T$. That is, for any (sensitivity) matrix $g'$ which for some orthogonal matrix $U'$ satisfies
\[
\delta \theta = U'(g')^T \delta S + O(\|\delta S\|^2),
\]
the sensitivity matrix $g$ satisfies the following inequalities for all $\ell < k$,
\[
\begin{align*}
\sum_{i \leq \ell} \sum_{j} g_{ij}^2 \geq & \sum_{i \leq \ell} \sum_{j} g'_{ij}^2 \\
\text{and} \\
\sum_{i \geq \ell} \sum_{j} g_{ij}^2 \leq & \sum_{i \geq \ell} \sum_{j} g'_{ij}^2
\end{align*}
\]
i.e. among all such sensitivity matrices $g$ squeezes as much of the sensitivity effect as possible into the lower $i$ components.

For the above reasons, we call $g_{ij}$, for $j = 1, \ldots, k$, the principal coefficients of sensitivity of $\text{MVN}(\mu(S), \Sigma(S))$ to changes in the $i$th component of the signal $S_i$, $i = 1, \ldots, k$.

2.5.3 The multiplexing capacities of a model

So far as modelling is concerned the typical situation is where the signals $S_i$ change some of the parameters $\theta_j$, $j = 1, \ldots, m$ of the model so that $\theta = \theta(S)$. Assume that the parameters that are changed by the signals have been reordered so that $\nu_1^2 = 2 \cdot \text{MX}(\theta_1, \ldots, \theta_j)$ satisfy $\nu_1 \geq \nu_2 \geq \cdots \geq \nu_s$ as can be done using the results of Sect. 2.4.1. Assume also that the signals have also been reordered so that $\nu_j$, where $\nu_j^2 = 2 \cdot \text{MX}(S_1, \ldots, S_j)$, similarly decrease. We show that then
\[
\prod_{i=1}^s \nu_j = (\det d\theta)^2 \prod_{i=1}^s \nu_j.
\]
where $d\theta = d\theta(S)$ is the derivative of $\theta$ with respect to $S$ at $S \in S$ as in Sect. 2.4.1. It follows from (15) that if these parameters have a large multiplexing capacity then signals for which $d\theta$ is well-conditioned will also multiplex well.

To show eq. (15) we note that in this case the FIM of the parameters $I_\theta$ and signals $I(S)$ are related by equation (11). Consider the corresponding sensitivity matrices $g_\theta$ and $g(S)$, respectively. By equation (11)
\[
(g(S))^T g(S) = d\theta^T I_\theta d\theta = (g_\theta d\theta)^T (g_\theta d\theta).
\]
Therefore, the matrices $g(S)$ and $g_\theta$ have the same singular values and the same determinant.

Let $QR = P g_\theta$ and $Q' R' = P' g(S)$ be the $QR$ decompositions from Sect. 2.4.1 Then $\nu_k = \text{MX}(\theta_1, \ldots, \theta_k) = R_{kk}^2/2$, $k = 1, \ldots, s$ and $\nu'_k = \text{MX}(S_1, \ldots, S_k) = R_{kk}'^2/2$, $k = 1, \ldots, s$. Thus,
\[
\prod_{i=1}^s \nu'_j / \prod_{i=1}^s \nu_j = \prod_{i=1}^s R_{jj}'^2 / \prod_{i=1}^s R_{jj}^2 = \det g(S)^2 / \det g_\theta^2 = (\det d\theta)^2
\]
3 Reaction Networks & Stochastic NF-κB model

Stochastic models of cellular processes in signaling and regulatory systems are usually described in terms of reaction networks. A system of multiple different molecular subpopulations has state vector, \( \mathbf{Y}(t) = (Y_1(t), \ldots, Y_n(t)) \) where \( Y_i(t), i = 1, \ldots, n, \) denotes the number of molecules of each species at time \( t \). These molecules undergo a number of possible reactions (e.g. transcription, translation, degradation) where the reaction of index \( j \) changes \( \mathbf{Y}(t) \) to \( \mathbf{Y}(t) + \mathbf{v}_j, \mathbf{v}_j \in \mathbb{R}^n \). The vectors \( \mathbf{v}_j \) are called stoichiometric vectors. Each reaction occurs randomly at a rate \( w_j(\mathbf{Y}(t)) \) (often called the intensity of the reaction), which is a function of \( \mathbf{Y}(t) \) but may also depend on \( t \).

Therefore the state of the system at some time \( t \) is
\[
\mathbf{Y}(t) = \mathbf{Y}(0) + \sum_j \mathbf{v}_j Z_j \left( \int_0^t w_j(\mathbf{Y}(s)) \, ds \right),
\]
with \( Z_j \) independent unit Poisson processes corresponding to the \( j \)-th reaction channel. The Stochastic Simulation algorithm (SSA) \[4\] use the properties of the above stochastic process to generate stochastic trajectories. These have exactly the same distribution with \( \mathbf{Y}(t) \). More specifically, the SSA, often referred as Gillespie algorithm, generates the time to the first next reaction, \( \Delta t \), at a given time \( t \) that has an exponential distribution with rate \( w_0(\mathbf{Y}(t)) = \sum_j w_j(\mathbf{Y}(t)) \) and the type of the next reaction that is of type \( j' \) with probability \( w_{j'}(\mathbf{Y}(t))/w_0(\mathbf{Y}(t)) \). Then the time \( t \) is updated to \( t + \Delta t \) and the state is updated to become \( \mathbf{Y}(t) + \mathbf{v}_{j'} \). The same steps are repeated until \( t > T \), for some time \( T > 0 \).

3.1 System size parameter

It is common in studying stochastic systems to introduce a system size \( \Omega \) which is a parameter that occurs in the intensities of the reactions \( w_j(\mathbf{Y}(t)) \). The precise description of this parameter depends on the system. In population models it might be considered to be of the same order of magnitude as the total population size while in chemical systems and cellular biology a natural choice is to use molar concentrations and therefore regard \( \Omega \) as Avogadro’s number in the appropriate molar units (e.g. nM\(^{-1}\)) multiplied by the volume of the reacting solution (e.g. the cell) in appropriate units (e.g. in litres (L)). In the NF-κB system, we consider it has units L/\(\mu\text{M}\).

The system size governs the size of the state fluctuations and therefore the size of the jumps. Larger system sizes generally imply less fluctuations and vice versa. In a certain sense the system size parameter is just a mathematical convenience to enable the study of the dependence of stochastic fluctuations upon system size. Indeed, the methods we develop in this paper can and should be applied to systems that do not involve a system size parameter but then it will be necessary to ensure that the population sizes achieved in the given system are large enough.

While having a system size parameter is not necessary to apply our methods, it allows one to study the dependence of stochastic fluctuations upon system size and to calculate the deterministic equations that describe the evolution of the concentration vector \( \mathbf{X}(t) = \mathbf{Y}(t)/\Omega \) in the limit of \( \Omega \to \infty \) (see next section). A sufficient condition to derive this limit is that the rates \( w_j(\mathbf{Y}(t)) \) depend upon \( \Omega \) (cf. \[8, 9, 10\]) as
\[
w_j(\mathbf{Y}) = \Omega w_j(\mathbf{Y}/\Omega).
\]
where \( w_j(x) \) the macroscopic (\( \Omega \to \infty \)) rates that generally depend on the concentration vector \( x = x(t) \). For example, linear degradation rates have the form \( w_{\text{deg}}(N) = cN \), where \( N \) stands for the relevant population number. The rate \( w_{\text{deg}}(N) \) is equal to \( \Omega u_{\text{deg}}(n) \), where \( u_{\text{deg}}(n) = cn \) the macroscopic rate and \( n = N/\Omega \) the corresponding concentration level. Similarly binding reaction rates have the form \( w_{\text{bind}}(N_1, N_2) = cN_1N_2 = \Omega u_{\text{bind}}(n_1, n_2) = \tilde{c}n_1n_2 \) where \( \tilde{c} = c\Omega \), and \( n_1 = N_1/\Omega, \ n_2 = N_2/\Omega \).
3.2 The classical deterministic approximation

Assuming condition (17), the equation (16) describing the time evolution of \( Y(t) \) can be written in terms of \( X = Y/\Omega \) as

\[
X(t) = X(0) + \sum_j \nu_j \Omega^{-1} Z_j \left( \int_0^t \Omega u_j(X(s)) ds \right).
\] (18)

Using the law of large numbers, as \( \Omega \to \infty \), we have that

\[
x(t) = x(0) + \sum_j \nu_j \int_0^t u_j(x(s)) ds,
\]

where here \( x(t) \) is the limit of convergence in probability for \( X(t) \) as \( \Omega \to \infty \). Consequently, \( x(t) \) satisfies the ordinary differential equation (ODE)

\[
\dot{x} = F(x(t)), \quad F(x) = \sum_j \nu_j u_j(x),
\] (19)

which is the classical macroscopic-deterministic approximation of the time-evolution of the concentration levels of the system. In matrix form, \( F(x) = Mu(x) \) where \( M = [\nu_1 \nu_2 \ldots \nu_r] \) the \( n \times r \) stoichiometry matrix and \( u(x(t)) = (u_1(x(t)), u_2(x(t)), \ldots, u_r(x(t)))^T \) the \( r \times 1 \) macroscopic rates vector. Herein we use \( \dot{x} \) to denote the time-derivative \( dx/dt \).

Throughout we consider autonomous systems (i.e. \( F(t) = F(x(t)) \)), but the method applies to non-autonomous systems with appropriate simple re-parameterisation (see for example [6]).

3.3 Stochastic NF-κB model

The model of NF-κB system used in this SI describes the oscillatory response of the system following stimulation by tumor necrosis factor alpha (TNFα). The system is initiated at its stable equilibrium state reached under the parameter settings of no stimulation (see Table B). This is followed by a period of TNFα stimulation either continuous or in the form of repeated pulses (e.g. 5 min stimulation every 100 minutes). The stimulation cause a transient oscillation, which in the case of pulses is repeated after every pulse, and in the case of continuous stimulation relaxes to a stable limit cycle.

The model of NF-κB considered here is a slight modification of that in [2]. In particular, the redundant (no feedback) variable of the phosphorylated IκBα is removed as well as the phosphorylated complex of NF-κB and IκBα and the inactive IKK that are also redundant because the sum of IKK and NF-κB are fixed in the model in [2]. These changes ensure that there are no rank-deficiencies in the covariance matrices of the LNA of the system.

We also write the system in a form where concentrations are in terms of the same volume. The original system in [2] is written in cytoplasmic concentrations for all species except the nuclear NF-κB, IκBα and their nuclear complex that are written in nuclear concentrations and are set to be 3.3 larger than cytoplasmic concentrations. Here, the necessary adjustments to the system equations are made so that all concentrations are of the same scale and no further conversions are necessary.

The ODE model for the NF-κB system is therefore given in Table C. The solution of the ODE system is provided in Fig A. The reactions and their rates used for the SSA are provided in Table D. The parameter values used in our simulations are provided in Table E.
The ODE system of the base NF-κB model used here which is the same as in [2] subject to some modifications explained in Section 3.3.

Table C: The ODE system of the base NF-κB model used here which is the same as in [2] subject to some modifications explained in Section 3.3.
| reaction                                                                 | rate                                                                 |
|-------------------------------------------------------------------------|----------------------------------------------------------------------|
| $N_c + I_c \xrightarrow{k_{1a}} NI_c$                                  | $k_{1a} I_c N_c / \Omega$                                           |
| $NI_c \xrightarrow{k_{d1a}} N_c + I_c$                                 | $k_{d1a} NI_c$                                                      |
| $N_n + I_n \xrightarrow{k_{1a}} NI_n$                                 | $k_{1a} I_n N_n / \Omega$                                           |
| $NI_n \xrightarrow{k_{d1a}} N_n + I_n$                                 | $k_{d1a} NI_n$                                                      |
| $K_a + I_c \xrightarrow{k_{2a}} K_a + I_p$                             | $k_{2a} K_a I_p$                                                    |
| $K_a + NI_c \xrightarrow{k_{2a}} NI_c$                                 | $k_{2a} K_a NI_c$                                                  |
| $NI_p \xrightarrow{k_{2a}} N_c$                                       | $k_{2a} NI_p$                                                      |
| $N_c \xrightarrow{k_{11}} N_n$                                        | $k_{11} N_c$                                                      |
| $N_n \xrightarrow{k_{e1}} N_c$                                        | $k_{e1} N_n$                                                      |
| $NI_n \xrightarrow{k_{e2a}} NI_c$                                     | $k_{e2a} N_n$                                                      |
| $I_c \xrightarrow{k_{i3a}} I_n$                                       | $k_{i3a} I_c$                                                      |
| $I_n \xrightarrow{k_{e3a}} I_c$                                       | $k_{e3a} I_n$                                                      |
| $\emptyset \xrightarrow{H} I_m$                                       | $(c_{1a} \Omega)^{N_h N_n} / (N_h + (\Omega / k_v)^n}$             |
| $I_m \xrightarrow{c_{2a}} I_m + I_c$                                   | $c_{2a} I_m$                                                      |
| $I_m \xrightarrow{c_{3a}} \emptyset$                                  | $c_{3a} I_m$                                                      |
| $I_c \xrightarrow{c_{4a}} \emptyset$                                  | $c_{4a} I_c$                                                      |
| $I_n \xrightarrow{c_{4a}} \emptyset$                                  | $c_{4a} I_n$                                                      |
| $NI_c \xrightarrow{c_{5a}} N_c$                                       | $c_{5a} NI_c$                                                      |
| $\emptyset \xrightarrow{H_A} A_m$                                     | $(c_{1a} \Omega)^{N_h N_n} / (N_h + (\Omega / k_v)^n}$             |
| $A_m \xrightarrow{c_{2a}} A_m + A$                                     | $c_2 A_m$                                                        |
| $A_m \xrightarrow{c_{3a}} \emptyset$                                  | $c_3 A_m$                                                        |
| $A \xrightarrow{c_{4a}} \emptyset$                                    | $c_4 A$                                                        |
| $K_i \xrightarrow{M_A} K_n$                                           | $k_p (TIKK - K_n - K_a) \Omega (k_{b20}/\Omega + A \times TNF\alpha)$ |
| $K_n \xrightarrow{TNFA \times K_2} K_a$                               | $TNF\alpha \times k_a K_n$                                        |
| $K_a \xrightarrow{k_i} K_i$                                           | $k_i K_a$                                                        |

Table D: Reactions of the base NF-κB system and their rates.
| parameter | description                                           | value            | measurement unit |
|-----------|-------------------------------------------------------|------------------|------------------|
| $k_p$     | IKKn production                                       | 0.0006           | $s^{-1}$         |
| $k_a$     | Activation caused by TNFa                             | 0.004            | $s^{-1}$         |
| $k_i$     | Spontaneous IKK activation                            | 0.003            | $s^{-1}$         |
| $k_{a1\alpha}$ | NFkB-IkB association                               | 0.5              | $\mu M^{-1}s^{-1}$ |
| $k_{d1\alpha}$ | NFkB-IkB dissociation                               | 0.0005           | $s^{-1}$         |
| $k_{c1\alpha}$ | Catalysis of IKK-IkB dimer                        | 0.074            | $s^{-1}$         |
| $k_{c2\alpha}$ | Catalysis of IKK-IkB-NFkB trimer                   | 0.37             | $s^{-1}$         |
| $k_{f2\alpha}$ | degradation of IkBa (IKK dependent from trimer)   | 0.1              | $s^{-1}$         |
| $c_{4\alpha}$ | Free IkBa degradation                                  | 0.0005           | $s^{-1}$         |
| $c_{5\alpha}$ | NFkB complexed IkBa degradation                        | 0.000022         | $s^{-1}$         |
| $k_{i1}$     | NFkB nuclear import                                   | 0.0026           | $s^{-1}$         |
| $k_{e1}$     | NFkB nuclear export                                   | 0.000052         | $s^{-1}$         |
| $k_{c2\alpha}$ | NFkB-IkB nuclear export                              | 0.01             | $s^{-1}$         |
| $k_{c3\alpha}$ | IkBa nuclear export                                   | 0.000335         | $s^{-1}$         |
| $h$         | Order of hill function                                | 2                | –                |
| $k$         | Hill constant                                         | 0.0650           | $\mu M/L$       |
| $c_{1\alpha}$ | IkBa mRNA synthesis                                   | 1.400e-07        | $\mu M^{-1}s^{-1}$ |
| $c_{2\alpha}$ | IkBa translation rate                                 | 0.5              | $s^{-1}$         |
| $c_{3\alpha}$ | IkBa mRNA degradation                                 | 0.003            | $s^{-1}$         |
| $c_{1}$     | IkBa mRNA synthesis                                   | 1.4e-07          | $\mu M^{-1}s^{-1}$ |
| $c_{2}$     | A20 mRNA translation                                  | 0.5              | $s^{-1}$         |
| $c_{3}$     | A20 mRNA degradation                                  | 0.00048          | $s^{-1}$         |
| $c_{4}$     | A20 degradation                                       | 0.0045           | $s^{-1}$         |
| $k_{5A20}$  | Half-max A20 inhibition concentration                 | 0.0018           | $\mu M/L$       |
| TNFα       | Tumor necrosis factor alpha level                     | 10               | ng/mL            |
| TNF-κB     | total NF-κB concentration                             | 0.08             | $\mu M$         |
| TIKK       | total IKK concentration                               | 0.08             | $\mu M$         |

Table E: The parameters of the base NF-κB system and the values used to derive their ODE solution.
3.4 The mNF-κB model

In addition to the variables and parameters in the base model, the mNF-κB model has the extra variables and parameters provided in tables $F$ and $K$, respectively. As you can see in table $G$ which includes the ODE system of the mNF-κB model, the rest of the system is unchanged.

| name     | description                                           | initial value |
|----------|-------------------------------------------------------|---------------|
| 1 $N_c$  | Free Cytoplasmic NF-κB                               | 0.0036        |
| 2 $I_c$  | Free Cytoplasmic IκBα                                 | 0.016         |
| 3 $(N-I)_c$ | Cytoplasmic complex of NF-κB and IκBα               | 0.072         |
| 4 $N_n$  | Free nuclear NF-κB                                   | 0.0042        |
| 5 $I_n$  | Free nuclear IκBα                                    | 0.0013        |
| 6 $(N-I)_n$ | Nuclear complex NF-κB and IκBα                    | 2.62e-4       |
| 7 $I_m$  | IκBα mRNA                                             | 2.04e-05      |
| 8 $K_n$  | IKK in neutral state                                 | 0.08          |
| 9 $K_a$  | IKK in active state                                  | 2.22e-16      |
| 10 $A_m$ | A20 mRNA                                              | 1.27e-05      |
| 11 $A$   | A20                                                   | 0.0014        |
| 12 $N_m$ | Free Cytoplasmic NF-κB                               | 0.0018        |
| 13 $N_m$I | Cytoplasmic complex of modified NF-κB and phosphorylated IκBα | 0.0360 |
| 14 $N_m$-I_p | Cytoplasmic complex of modified NF-κB and phosphorylated IκBα | 0       |
| 15 $(N_m)_n$ | Free nuclear modified NF-κB                         | 0.0021        |
| 16 $(N_m$I)p_n | Free complex of modified NF-κB and phosphorylated IκBα | 0.0001 |

Table $F$: The variables of mNF-κB model and the initial conditions (in $\mu$M, concentration per cell) used to derive the ODE solution. The initial conditions are set to the equilibrium fixed point derived under no stimulation. The extra variables, compared to the base model, are marked in blue.
\[
\begin{align*}
\dot{N}_c &= k_{d1a}(N-I)_c - k_{a1a}N_cI_c - k_{i1}N_c + c_{5a}(N-I)_c + k_vk_e1N_n + k_02a(N-TNF) - p_{a1}S_2 \frac{N_c}{N_c + k_{pmf}} + p_{d1} \frac{N_m}{N_m + k_{mnf}} \\
\dot{I}_c &= k_{d1a}(N-I)_c - k_{a1a}N_cI_c - k_{i3a}I_c + k_vk_e3aI_n - c_{4a}I_c + c_{2a}I_m - k_{c1a}K_aI_c \\
(\dot{N}-I)_c &= k_{a1a}N_cI_c - k_{d1a}(N-I)_c + k_vk_e2a(N-I)_n - c_{5a}(N-I)_c - k_{c2a}K_a(N-I)_c \\
\dot{N}_n &= k_{d1a}(N-I)_n - k_vk_a1aN_nI_n + k_{i1}N_c - k_vk_e1N_n \\
\dot{I}_n &= k_{d1a}(N-I)_n - k_vk_a1aN_nI_n + k_{i3a}I_n - k_vk_e3aI_n - c_{4a}I_n \\
(\dot{N}-I)_n &= k_vk_a1a(N-I)_n - k_{d1a}(N-I)_n - k_vk_e2a(N-I)_n \\
\dot{I}_m &= c_1a\left(\frac{N_n+(N_m)n}{N_n+(N_m)n}\right)^{\beta} - c_3A_m \\
\dot{K}_n &= k_p(TIKK - K_n - K_a)\left(k_{bA20}/(k_{bA20} + A \times TNFa)\right) - k_aTNFaK_n \\
\dot{A}_m &= c_1a\left(\frac{N_n+(N_m)n}{N_n+(N_m)n}\right)^{\beta} - c_2A_m - c_4A \\
\dot{N}_m &= k_{d1a}(N_m-I_n) - k_{a1a}N_mI_c - k_{i1}N_m + k_vk_e1(N_m)_n + c_{5a}N_m-I + k_{i2a}(N_m-I_p) + p_{a1}S_2 \frac{N_c}{N_c + k_{pmf}} - p_{d1} \frac{N_m}{N_m + k_{mnf}} \\
(\dot{N}_m-I) &= k_{a1a}N_mI_c - k_{d1a}N_m-I + k_vk_e2aN_m-I_n - c_{5a}N_m-I - k_{c2a}K_aN_m-I \\
(\dot{N}_m-I_p) &= k_{c2a}K_aN_m-I - k_{i2a}N_m-I_p \\
(\dot{N}_m)_n &= k_{d1a}N_mI_n - k_vk_a1a(N_m)_nI_n + k_{i1}N_m - k_vk_e1(N_m)_n \\
(\dot{N}_m)_n &= k_vk_a1a(N_m)_nI_n - k_{d1a}N_m-I_n - k_vk_e2aN_m-I_n
\end{align*}
\]

**Table G:** The ODE system for the mNF-κB model. This is an extension of the model provided in [2] to include modification of NF-κB (see also Section 3.3). The extra equations, compared to the base model, are marked in blue.

| parameter | description                  | value | measurement | unit |
|-----------|-----------------------------|-------|-------------|------|
| \(S_2\)  | signal controlling NF-κB modification | 1     |             | s\(^{-1}\) |
| \(p_{a1}\) | NF-κB modification rate       | 0.0175|             | s\(^{-1}\) |
| \(p_{d1}\) | NF-κB modification rate       | 0.0175|             | s\(^{-1}\) |
| \(k_{pmf}\) | half-max NF-κB modification level | 0.01  | \(\mu M/L\) |     |
| \(k_{mnf}\) | half-max NF-κB un-modification level | 0.01  | \(\mu M/L\) |     |

**Table H:** The extra parameters of mNF-κB system, compared to the base model, and the values used to derive their ODE solution.
3.5 Another model for NF-κB modification

We now consider another model for NF-κB modification where the NF-κB molecules that are bound by IκBα molecules, are transformed to another complex of modified NF-κB molecules bounded by IκBα molecules. For the modification to happen, a transition to an intermediate state needs to happen first. This transition is controlled by an independent signal $S_2$. A similar mechanism using an intermediate state is used for the reverse modification. The diagram of the model, which we will call amNF-κB model, is depicted in Fig D(a), while the solutions of the ODE equations provided in table J are provided in Fig D. As we can see in Fig D(b), the singular values of the FIM of amNF-κB are larger than the singular values of the base and mNF-κB models. The multiplexing capacities are also increased with parameters related to this modification taking some of the higher values (see Fig 3.5(c)). The sensitivity coefficients are generally increased with those parameters related to the modification having high sensitivity (see Fig 3.5(d)).
Figure C: Diagram of the main reactions of the amNF-κB model described in SI section 3.5.

Table I: The variables of the amNF-κB model described in section 3.5 and the initial conditions (in µM, concentration per cell) used to derive the ODE solution. The initial conditions are set to the equilibrium fixed point derived under no stimulation. The extra variables, compared to the base model, are marked in blue.
Table J: The ODE system for the amNF-κB model described in section 3.5. The extra equations, compared to the base model, are marked in blue.
Table K: The extra parameters of amNF-κB model described in section 3.5, compared to the base model, and the values used to derive their ODE solution.

| parameter | value | measurement unit |
|-----------|-------|------------------|
| TM        | 0.0008 | µM/L             |
| TR        | 0.0004 | µM/L             |
| p_a1      | 6     | s⁻¹              |
| p_d1      | 0.0035 | s⁻¹              |
| p_k1      | 0.0035 | s⁻¹              |
| p_a2      | 6     | s⁻¹              |
| p_d2      | 0.0035 | s⁻¹              |
| p_k2      | 0.0035 | s⁻¹              |
| S2        | 0.62  |                  |

Figure D: The solution of the variables of the ODE system of the amNF-κB model described in section 3.5.
Figure E: (a) The singular values $\sigma_i$ of the amNF-κB, mNF-κB and base model; (b) The multiplexing capacities $v_j$ of the amNF-κB model. (c) The principal sensitivity coefficients of the amNF-κB model. Larger values indicate higher sensitivity of the amNF-κB model to changes in the value of the corresponding parameter.

4 LNA, pcLNA and calculating the FIM

We next describe the LNA \cite{1} and pcLNA \cite{13} as well as how to derive the FIM using pcLNA.

4.1 The linear noise approximation (LNA)

The LNA ansatz describes the state $X(t)$ at some time $t$ as the sum of the deterministic solution $x(t)$ of \cite{19} and the noise, $\xi(t)$, scaled by $\Omega^{-1/2}$, i.e.

$$X(t) = x(t) + \Omega^{-1/2}\xi(t). \quad (20)$$

The noise satisfies the linear stochastic differential equation (sde)

$$d\xi(t) = J(x(t))\xi(t)dt + \sqrt{MU(x(t))M^T}dW_t \quad (21)$$

and $W_t$ an $n \times 1$ Wiener process (i.e. $W_t$ is a random vector of independent one-dimensional Wiener processes). Here $J(x(t))$ is the Jacobian matrix of \cite{19}, i.e.

$$J(x) = \left(\frac{dF_i}{dx_j}\right)_{ij}, \quad (22)$$

$M$ the stoichiometry matrix, and $U(x)$ the diagonal matrix with main diagonal the rates vector $u(x)$. The noise sde can be solved and the solution can be written as

$$\xi(t) = C(t_0, t)\xi(t_0) + \eta(t_0, t), \quad \eta(t_0, t) \sim \text{MVN}(0, V(t_0, t)), \quad (23)$$

where $C(t_0, t)$ is the fundamental matrix of \cite{19}, i.e. $C(t_0, t)$ is the solution of the initial value problem

$$\dot{C} = JC, \quad C(t_0, t_0) = I, \quad (24)$$

and the symmetric positive-definite matrix $V(t_0, t)$ is the solution of the initial value problem

$$\dot{V}(t_0, t) = JV + VJ^T + MUM^T, \quad V(t_0, t_0) = 0. \quad (25)$$
Equation 23 implies that if \( \xi(t_0) \sim MVN(m_0, V_0) \), then \( P_{LNA}(\xi(t)|\xi(t_0)) \sim MVN(m_t, V_t) \), with

\[
m_t = C(t_0, t)m_0, \quad V_t = C(t_0, t)V_0C(t_0, t)^T + V(t_0, t)
\]

and the LNA anstanz in (20) that \( P_{LNA}(X(t)|\xi(t_0) \sim MVN(m_0, V_0)) \) is \( MVN(\mu_t, \Sigma_t) \), where

\[
\mu_t = x(t) + \Omega^{-1/2}m_t, \quad \Sigma_t = \Omega^{-1}V_t.
\]

It is worth noting that the LNA can be derived by applying the Central Limit Theorem, as \( \Omega \to \infty \), to the Poisson random variables in (18) (see [1], [8], [9]). It was also derived by van Kampen (see [10]) as an expansion, in powers of \( \Omega \), of Kolmogorov’s forward equations (also called Master equation) of the Poisson process \( Y(t) \) in [16] ignoring terms of order \( O(\Omega^{-1/2}) \).

4.2 Phase corrected LNA (pcLNA)

The pcLNA method described below is very similar to this in [13]. However, for clarity, we provide a description more appropriate to the current context. Unlike [13], we consider solutions \( x = g(t) \), \( t \geq 0 \), of (19) without assuming that \( \gamma \) is a limit cycle. However, especially for low \( \Omega \), that is where the fluctuations are large, the method provides better approximation when the underlying deterministic solutions are such that transversal sections are well defined (e.g. oscillatory systems).

We first provide several definitions relevant to the theory developed herein and in I.

**Transversal sections** Consider the solution \( \gamma \) of the ode in [19] given by \( x = g(t) \), \( t > 0 \). By a transversal section through \( \gamma \) we mean an \((n - 1)\)-dimensional linear hyperplane \( S_x \) containing \( x \) and transversal to the tangent vector, \( F(x) \), to \( \gamma \) at \( x \). A particular example is the hyperplane normal to \( \gamma \) at \( x \). A transversal system is a family \( S_{g(t)} \) of transversal sections that vary smoothly with \( t \) in the sense that the unit normal vector to \( S_{g(t)} \) varies smoothly with \( t \).

A transversal system defines a mapping \( G \) of a neighbourhood of \( \gamma \) onto \( \gamma \) where if \( X \in S_x \) then \( G(X) = x \in \gamma \). In cases where \( X = X(t) \) lies in more than one transversal sections, then \( G(X(t)) = x(s) \) with \( s = \arg \min_{s': G(X(t)) = x(s')} |s' - t| \) the closest time to \( t \). We denote this mapping for the normal transversal system by \( G_N \).

**Computation of the normal mapping \( G_N(X(t)) \)** To compute the mapping \( G_N(X) = x \in \gamma \), of \( X = X(t) \), for a given solution \( x = g(t) \), \( t > 0 \), of the ode in [19], we wish to minimise the squared distance

\[
\|X - g(s)\|^2
\]

with respect to \( s > 0 \). Note that the minimum \( s \) is achieved when

\[
h(s) = F(g(s))^T(X - g(s)) = 0,
\]

i.e. when the tangent vector \( F(g(s)) \) is orthogonal to \( X - g(s) \) and thus \( X \) lies on the normal transversal section \( S_{g(s)} \). To find the appropriate \( s \), we first use a Newton’s optimisation to derive the approximate limit of the sequence

\[
s_{n+1} = s_n - h(s)/\dot{h}(s), \quad n = 0, 1, \ldots
\]

for different \( s_0 \), stopping the search when the square of the step \( h(s)/\dot{h}(s) \) is appropriately small. If different limits are achieved for different \( s_0 \), we choose the one that is closest to \( t \). Note that other methods of optimisation could be used here.
Adapted coordinate systems An adapted coordinate system $C_{g(t)}$ at a point $g(t)$ on $\gamma$ is determined by a set of orthonormal basis vectors $e_1(t), \ldots, e_n(t)$ with $e_1(t)$ the unit normal vector to $S_{g(t)}$ and the vectors $e_2(t), \ldots, e_n(t)$ forming an orthonormal basis of $S_{g(t)}$. If these are defined for $t$ in some interval in $\mathbb{R}^+$ then we always assume that the $e_i(t)$ have smooth (i.e. $C^2$) dependence upon $t$. It is important that the coordinates are defined by an orthonormal basis in the original coordinates because this effectively preserves the covariance matrix in the sense that a covariance matrix $V$ in the adapted coordinates is $RV\,R^T$ in the original coordinates with $R$ a real orthogonal matrix. In particular, the eigenvalues are preserved.

The state $X$ of the process (written in original coordinates corresponding to concentrations of each reacting molecule) relates to the adapted coordinates $(X^{(1)}, X^{(2)})$ at $x = g(t)$ with $X^{(2)} \in S_{g(t)}$ through the orthogonal matrix $R_t = [R^{(1)}_t \quad R^{(2)}_t]$ by

$$
(X^{(1)}, X^{(2)})^T = R_t(X - x(t)) \quad \text{and} \quad X - x(t) = R^{(1)}_t X^{(1)} + R^{(2)}_t X^{(2)}.
$$

Transversal points Consider a stochastic trajectory $X(t_i)$, $i = 0, 1, 2, \ldots$ and let $g(t)$, $t \geq 0$, be a solution of (19) with $g(t_0) = g_0$. Suppose that the stochastic trajectory has initial condition $X(0)$ in the neighbourhood of $g_0$. For the SSA trajectories presented in I, this is satisfied by taking $X_0 = g_0$, but this is not necessary. Find the transversal section $S_{g(s)}$ for which $G_N(X(t_0)) = g(s)$ and $S_0 = \arg\min_{s', G_N(X(t_0)) = g(s')} |s' - t_0|$. Then, for $i = 1, 2, \ldots$, define the transversal sections $S_{g(s_i)}$, $i = 1, 2, \ldots$, such that

$$
G_N(X(t_i)) = g(s_i) \quad \text{and} \quad s_i = \arg\min_{s'} |s' - s_{i-1} + (t_i - t_{i-1})|.
$$

Suppose that we wish to find the point, $Q_x$, at which the stochastic trajectory intersects the transversal section $S_{x=g(s)}$ where $s \in (s_i, s_{i+1})$ for some $i > 0$. Then some form of interpolation between $X(t_i)$ and $X(t_{i+1})$ needs to be used. For the SSA trajectories presented in I, we first use that there are no jumps between $t_i$ and $t_{i+1}$, then write $X(t_i)$ in coordinates $(r_i, q_i)$ adapted at $g(s)$, i.e. use the $C_{g(s)}$ coordinates, and take the transversal point $Q_x = q_i$.

Note that backward moves and therefore multiple intersections to the same transversal section are possible. However, as we have shown in [13], the distribution of the number of intersections decreases exponentially and the distribution of the transversal points in different intersection is indistinguishable. Therefore, hereafter, we assume that in each trajectory there is only a single intersection to each transversal section, $S_x$, of $x \in \gamma$, called transversal point $Q_x$.

4.2.1 The pcLNA distribution

Consider a solution $g(t)$, $t > 0$ of (19) that satisfies $g_0 = g(0)$ and a point $x = g(t)$ of this solution. Let $S_{x=g(t)}$ be the corresponding transversal section at $x = g(t)$. We wish to approximate the distribution of the transversal point $Q_{x=g(t)}$ as defined in section 4.2. We use the adapted coordinates $C_x$ and write the state of the system at time $t$ as $X(t) = (X^{(1)}(t), X^{(2)}(t))^T$, with $X^{(2)}(t) \in S_x$. Then the transversal distribution $P(Q_x | X(0))$ is estimated in pcLNA by the distribution

$$
P_{\text{LNA}}(X^{(2)}(t) | X(0)).
$$

This is the LNA distribution of the state of the system on the transversal section $S_{x=g(t)}$ at time $t$.

To compute this distribution, henceforth called pcLNA distribution and written

$$
P_{\text{pcLNA}}(Q_x | X(0)),$$

suppose that $\xi(0) \sim \text{MVN}(m_0, S_0)$ and use the results in section 4.1 to compute $P(X(t) | \xi(0) \sim \text{MVN}(m_0, S_0))$ which is $\text{MVN}(\mu_t, \Sigma_t)$ with parameters as in [27] in
the original coordinates (that correspond to concentration levels of each of the reacting molecules). Then, find an orthogonal matrix \( R_x = [R_x^{(1)} \ R_x^{(2)}] \) such that \( R_x^T (X(t) - x) = (X^{(1)}(t), X^{(2)}(t))^T \) with \( X^{(2)}(t) \in S_{x=g(t)} \). The pcLNA distribution \( P_{pcLNA}(Q_x|x(0) \sim MVN(m_0, S_0)) \) is \( MVN(\mu_x, \Sigma_x) \), where

\[
\mu_x = \Omega^{-1/2} m_t^{(2)}, \quad \Sigma_x = \Omega^{-1} V_t^{(2,2)},
\]

and the moments of the noise process on the transversal are

\[
m_t^{(2)} = R_x^{(2)T} m_t, \quad V_t^{(2,2)} = R_x^{(2)T} V_t R_x^{(2)},
\]

with \( m_t, V_t \) given in [26]. This pcLNA distribution is compared with the empirical distribution of transversal points derived from SSA simulations as explained in the previous section. The results are presented in Figure 4 and demonstrate that the two distributions are hardly distinguishable.

Note that the above pcLNA distributions are slightly simpler than those presented in [13]. Here the marginal LNA distributions of \( X^{(2)} \) are used instead of the conditional distributions \( (X^{(2)}|X^{(1)}=0) \). The performance of the two distributions compared to the SSA simulations are very similar with differences identified only during the initial transient response of NF-\( \kappa \)B to TNF\( \alpha \) signals. The simpler marginal distributions provided here give slightly better fit during this period due to their extra variability that fits better the SSA simulations. The properties of pcLNA distributions, in terms of accuracy compared to SSA, robustness as \( t \to \infty \), and analytical tractability are thoroughly discussed in [13].

### 4.2.2 The pcLNA joint distribution of multiple transversals

We now wish to approximate the joint distribution of multiple transversal points \( P(Q_{1:k} | X(0) \sim MVN(\mu_0, \Sigma_0)) = P(Q_{x_1}, Q_{x_2}, \ldots, Q_{x_k} | X(0) \sim MVN(\mu_0, \Sigma_0)) \) where \( x_i = g(t_i) \) and \( 0 \leq t_1 \leq \cdots \leq t_k \). This can be written as a product of transition probabilities

\[
P(Q_{1:k} | X(0)) = P(Q_{x_1}, Q_{x_2}, \ldots, Q_{x_k} | X(0)) = P(Q_{x_1} | X(0)) \prod_{i=1}^{k-1} P(Q_{x_{i+1}} | Q_{x_i})
\]

The first transition probability on the rhs of the above equation, as explained in the previous section, is approximated by the \( MVN(\mu_{x_1}, \Sigma_{x_1}) \) distribution, by taking \( \mu_0 = x_0 + \Omega^{-1/2} m_0 \), \( \Sigma_0 = S_0/\Omega \) and \( x = x_1 = g(t_1) \). To derive the approximation of the transition probabilities from \( Q_{x_i} \) to \( Q_{x_{i+1}} \) note that

\[
\begin{align*}
X^{(2)}(t_{i+1}) &= R_x^{(2)T} (X(t_{i+1}) - x(t_{i+1})), \quad \text{by (28)} \\
&= R_x^{(2)T} \Omega^{-1/2} \xi(t_{i+1}), \quad \text{by (20)} \\
&= R_x^{(2)T} \Omega^{-1/2} (C(t_i, t_{i+1}) \xi(t_i) + \eta(t_i, t_{i+1})), \quad \text{by (23)} \\
&= \Omega^{-1/2} (R_x^{(2)T} \eta(t_i, t_{i+1}) + R_x^{(2)T} C(t_i, t_{i+1})(X(t_i) - x(t_i))), \quad \text{by (20)} \\
&= \Omega^{-1/2} (R_x^{(2)T} \eta(t_i, t_{i+1}) + R_x^{(2)T} C(t_i, t_{i+1})(R_x^{(1)} X^{(1)}(t_i) + R_x^{(2)} X^{(2)}(t_i))), \quad \text{by (28)}.
\end{align*}
\]

We assume that the term describing the transition through \( C(t_i, t_{i+1}) \) of the tangential fluctuations at time \( t_i, R_x^{(1)} X^{(1)}(t_i) \), to fluctuations on the transversal coordinate, \( R_x^{(2)} \), at time \( t_{i+1} \), i.e.

\[
\Omega^{-1/2} R_x^{(2)T} C(t_i, t_{i+1}) R_x^{(1)} X^{(1)}(t_i)
\]
is small compared to the transition of transversal fluctuations and ignore it to derive that
\( P_{\text{pcLNA}}(Q_{x_{i+1}} | Q_{x_i} = q_i) = P_{\text{LNA}}(X^{(2)}(t_{i+1}) | X^{(2)}(t_i) = q_i) \) is \( \text{MVN}(\mu_{x_{i+1}}, \Sigma_{x_{i+1}}) \), where

\[
\hat{\mu}_{x_{i+1}} = C^{(2,2)}(t_i, t_{i+1}) q_i, \quad C^{(2,2)}(t_i, t_{i+1}) = R_{x_{i+1}}^{(2)} T C(t_i, t_{i+1}) R_{x_i}^{(2)},
\]

and

\[
\Sigma_{x_{i+1}} = \Omega^{-1} V^{(2,2)}(t_i, t_{i+1}), \quad V^{(2,2)}(t_i, t_{i+1}) = R_{x_{i+1}}^{(2)} T V(t_i, t_{i+1}) R_{x_{i+1}}^{(2)}.
\]

More generally, the pcLNA transition probability with \( X^{(2)}(t_i) \sim \text{MVN}(\mu_{x_i}, \Sigma_{x_i}) \),

\[
P_{\text{pcLNA}}(Q_{x_{i+1}} | Q_{x_i} \sim \text{MVN}(\mu_{x_i}, \Sigma_{x_i})) = P_{\text{LNA}}(X^{(2)}(t_{i+1}) | X^{(2)}(t_i) \sim \text{MVN}(\mu_{x_i}, \Sigma_{x_i}))
\]
is \( \text{MVN}(\mu_{x_{i+1}}, \Sigma_{x_{i+1}}) \) where (again ignoring the tangential fluctuations transduced on the transversal)

\[
\mu_{x_{i+1}} = C^{(2,2)}(t_i, t_{i+1}) \mu_{x_i},
\]

and

\[
\Sigma_{x_{i+1}} = \Sigma_{x_{i+1}} + C^{(2,2)}(t_i, t_{i+1}) \Sigma_{x_i} C^{(2,2)}(t_i, t_{i+1})^T.
\]

Then, the pcLNA joint distribution is

\[
P_{\text{pcLNA}}(Q_{x_2}, Q_{x_1} | X(0) \sim \text{MVN}(\mu_0, \Sigma_0)) \propto \exp \left( -\frac{1}{2} \left( (Q_{x_1} - \mu_{x_1})^T \Sigma_{x_1}^{-1} (Q_{x_1} - \mu_{x_1}) + (Q_{x_2} - C^{(2,2)}(t_1, t_2) Q_{x_1})^T \Sigma_{x_2} (Q_{x_2} - C^{(2,2)}(t_1, t_2) Q_{x_1}) \right) \right)
\]

and by adding and subtracting the term \( C^{(2,2)}(t_1, t_2) \mu_{x_1} \) in the second term of the rhs above, we have that this can be written as the kernel of a \( \text{MVN} \) distribution with mean vector

\[
(\mu_{x_1}, \mu_{x_2})
\]

and precision matrix (inverse of covariance matrix)

\[
\begin{pmatrix}
\Sigma_{x_1}^{-1} + C^{(2,2)}(t_1, t_2)^T \Sigma_{x_2}^{-1} C^{(2,2)}(t_1, t_2) & -C^{(2,2)^T}(t_1, t_2) \Sigma_{x_2}^{-1} \\
-C^{(2,2)}(t_1, t_2) \Sigma_{x_2}^{-1} C^{(2,2)^T}(t_1, t_2) & \Sigma_{x_2}^{-1}
\end{pmatrix}.
\]

By induction we can derive that the pcLNA joint distributions

\[
P_{\text{pcLNA}}(Q_{x_1}, Q_{x_2}, \ldots, Q_{x_k} | X(0) \sim \text{MVN}(\mu_0, \Sigma_0)), \text{ for } k = 2, \ldots, \]

with mean vector

\[
\mu_{x_{1:k}} = (\mu_{x_1}, \mu_{x_2}, \ldots, \mu_{x_k})
\]

and precision matrix, \( A_{x_{1:k}} \), with a block tridiagonal form where the diagonal blocks, \( A_{x_{1:k}}^{(ii)} \), are

\[
\Sigma_{x_i}^{-1} + (C^{(2,2)}(t_i, t_{i+1}))^T (\Sigma_{x_{i+1}}^{-1})^{-1} C^{(2,2)}(t_i, t_{i+1}), \text{ for } i = 1, 2, \ldots, k - 1
\]

and

\[
\Sigma_{x_i}^{-1}, \text{ for } i = k,
\]

upper diagonal entries, \( A_{x_{1:k}}^{(i,i+1)} \), \( i = 1, 2, \ldots, k - 1, \)

\[
-C^{(2,2)}(t_i, t_{i+1})^T \Sigma_{x_{i+1}}^{-1}
\]

and lower diagonal entries, \( A_{x_{1:k}}^{(i+1,i)} \), \( i = 1, 2, \ldots, k - 1, \)

\[
-\Sigma_{x_{i+1}}^{-1} C^{(2,2)}(t_i, t_{i+1}).
\]
To write the pcLNA joint distribution in the original coordinates we can use the transformation $x_{1:k} + R_{1:k}Q_{1:k}$ where $x_{1:k} = (x_1, x_2, \ldots, x_k)^T$, $R_{1:k} = \text{Diag}(R_{t_1}^{(2)}, \ldots, R_{t_k}^{(2)})$ to get the mean vector
\[ \mu_{1:k} = x_{1:k} + R_{1:k} \mu_{x_{1:k}} \]
and precision matrix
\[ A_{1:k} = R_{1:k} A_{x_{1:k}} R_{1:k}^T. \]

As discussed above, the pcLNA joint distributions have been shown to have hardly distinguishable distributions to the empirical transversal distributions derived by SSA in the NF-κB system (see Figure 4 in [1] and [13]).

### 4.3 pcLNA simulation

The approach is to amend the LNA Ansatz $X(t) = g(t) + \Omega^{-1/2} \xi(t)$ to $X(t) = g(s) + \Omega^{-1/2} \kappa(s)$ where $g(s) = G_N(X(t))$ and to use resetting of $t$ to $s$ to cope with the growth in the variance of phase deviations keeping the LNA fluctuation $\kappa(s)$ normal to $\gamma$. While for free-running oscillators the variance of $\xi(t)$ grows without bound as $t$ increases, $\kappa(s)$ has uniformly bounded variance.

The pcLNA simulation algorithm iteratively uses standard LNA steps of length $\Delta \tau$ to move from a state $X(s_{i-1})$ to a new state $X(s_{i-1} + \Delta \tau) = X_i$, $i = 1, 2, \ldots$. After each LNA step, the phase of the system is reset or “corrected” such that $g(s_i) = G_N(X_i)$ and the (global) fluctuations $\xi(s_{i-1} + \Delta \tau) = \Omega^{-1/2} (X_i - g(s_{i-1} + \Delta \tau))$ are replaced by the normally transversal fluctuation $\kappa(s_i) = \Omega^{-1/2} (X_i - g(s_i))$ which are MVN distributed and, as we showed in the previous section, approximate well the transversal fluctuations under the exact Markov Jump process.

The steps of the pcLNA simulation algorithm are described next in more detail (see also Fig 7).

1. Choose a time-step size $\Delta \tau > 0$.
2. Input initial conditions $\kappa(S_0)$ and $X_0 = g(S_0) + \Omega^{-1/2} \kappa(S_0)$.
3. For iteration $i = 1, 2, \ldots$
   
   - (a) sample $\xi(s_{i-1} + \Delta \tau)$ from $\text{MVN}(C_i \kappa(s_{i-1}), V_i)$;
   - (b) compute $X_i = g(s_{i-1} + \Delta \tau) + \Omega^{-1/2} \xi(s_{i-1} + \Delta \tau)$;
   - (c) set $s_i$ to be such that $G_N(X_i) = g(s_i)$ and $\kappa_i = \Omega^{1/2} (X_i - g(s_i))$.

In the for loop of step [3], $C_i = C(s_{i-1}, s_i + \Delta \tau)$ and $V_i = V(s_{i-1}, s_i + \Delta \tau)$ that are derived as solutions of [24] and [25]. The simulated sample $X_i$ corresponds to time $t_i = t_0 + i \Delta \tau$, $i = 1, 2, \ldots$, where $t_0$ is the initial time. The time $t_i$ is not necessarily equal to the phase $s_i$, defined by the relation $g(s_i) = G_N(X_i)$, which is stochastic and has variance linearly increasing with the time step $\Delta \tau$.

If one wants to record simulated trajectories at a finer time-scale than $\Delta \tau$ then one can run the algorithm with $\Delta \tau$ replaced by $\Delta \tau/M$ for some integer $M > 1$ and only carry out the phase correction in step 3(c) every $M$th step and at all the other steps just proceeding as in the standard LNA (ignoring step 3(c)). This gives the same distribution as if the intermediate points had not been calculated because of the transitive nature of the LNA i.e. the distribution $P_{LNA}(X(s + t)|X(0))$ is equal to the distribution $P_{LNA}(X(t)|X(s) \sim P_{LNA}(X(s)|X(0)))$. We have noticed that roughly 3-5 corrections per peak-to-peak time typically give good balance between speed and precision.
Figure F: The main steps in the pcLNA simulation algorithm. The solid horizontal bars below the horizontal axis are all of length $\Delta \tau$, the basic time step of the algorithm. The black arrows show $\hat{\xi}_i = \Omega^{-1/2} \xi(s_{i-1} + \Delta \tau)$ and the grey arrows $\hat{\kappa}_i = \Omega^{-1/2} \kappa(s_i)$. Having calculated $\hat{\kappa}(s_{i-1})$ one uses $\kappa(s_{i-1})$ as the initial state and updates it using the LNA and a time-step $\Delta \tau$ to obtain $\xi$ at $s_{i-1} + \Delta t$.

Then $\xi(s_{i-1} + \Delta \tau)$ is replaced by $\kappa(s_i)$ so that $g(s_{i-1} + \Delta \tau) + \Omega^{-1/2} \xi(s_{i-1} + \Delta \tau) = g(s_i) + \Omega^{-1/2} \kappa(s_i)$ where $\kappa(s_i)$ is normal to the limit cycle. Therefore, $s_i$ gives the phase of $\kappa(s_i)$ and the corresponding time is $t_i = t_0 + i \Delta \tau$.

### 4.4 Channel capacity estimation

In our comparisons to the experimental results of the NF-κB system response to TNFα stimulus in the section “Mutual Information” and Figure 4(d) of I, we allowed the total number of NF-κB molecules to vary in different cells-simulated trajectories. Specifically, at the beginning of the simulation of each trajectory, we first generated the total number of NF-κB molecules from a log-Normal distribution that is previously used in [16]. This has pdf

$$f(x) = \frac{A}{x \sigma \sqrt{2\pi}} \exp \left( -\frac{(\log(x) - \mu)^2}{2\sigma^2} \right), \quad x > 0$$

and the parameters

$$A = 10^5, \quad \sigma = 1/\sqrt{2}, \quad \mu = -1/4$$

are chosen to give a mean number of molecules equal to $10^5$ that agrees with the fixed value assumed in the deterministic model used in [2] as well as the rest of the models used here. The log-Normal is chosen to give a positive support (i.e. only positive number of the total number of molecules are possible), a single mode value and tunable variance.

After generating the number of NF-κB molecules, we used the pcLNA simulation algorithm to obtain a stochastic trajectory. We initiated the stochastic simulation from the equilibrium point under no dose and immediately applied a continuous TNFα dose. We then obtained the nuclear NF-κB concentration at $t = 30$ min.

The dose was assumed to have a discrete distribution with weights obtained using a constrained optimisation performed by the fmincon function in MATLAB [11].

We show the results assuming that the TNFα dose has only two levels (high and low), but the use of medium doses did not change the estimation substantially.

### 4.5 Channel capacity under the stochastic model of NF-κB in Tay et al. [16]

We used the code provided by [16] to run 500 simulations over a high (10ng/mL) and a low (0.5ng/mL) TNF dose. Similarly to above, we collected the nuclear concentration of NF-κB at $t = 30$ min after the initiation of the continuous TNFα stimulus. The histograms are provided in Fig G. The channel capacity was estimated to be 0.72 assuming that the dose has a discrete distribution. The optimal weights of the dose distribution were derived using a constrained optimisation performed by the fmincon function in MATLAB [11].
4.6 Simulation of EGR1 and COX-2 genes.

Here we explain the stochastic simulation of copies of mRNA molecules of the EGR1 and COX-2 genes as targets of the NF-κB. As explained in I, we use a similar model as the one described in SI section 3.4; the difference being that the modification is now promoted by the active state of IKK. More specifically, in equation 12 in table G the term \( p_{a1} S_2 N_{c+k_{pmf}} \) is replaced with \( p_{a1} K_a x S_2 N_{c+k_{pmf}} \) and the parameters \( p_{a1} = 0.0025, S_2 = 1, k_{pmf} = 0.01 \). The reverse modification parameters are also slightly changed to \( k_{pmf} = 0.01, p_{d1} = 1 \times 10^{-5} \).

Because the model does not include any feedback of these genes to the NF-κB system, they can be simulated separately to the base system, and this makes the hybrid algorithm explained below somewhat simpler than hybrid algorithms such as those discussed for example in [17]. We therefore used the pcLNA simulation algorithm to derive stochastic trajectories of the base system and then applied the SSA to generate stochastic simulations for the EGR1 and COX-2 genes. The reason for using SSA in the latter simulation rather than LNA or pcLNA is that EGR1 and COX-2 activation is highly stochastic and not oscillatory. This method proved to be fairly efficient for the purposes of our demonstration, as we exploited the computational advantages of using pcLNA in the base system, while SSA simulations of only two genes expressed in similar scales proved to also be fairly efficient.

The algorithm proceeds as follows. Let \( Y^F = (Y^F(0), Y^F(t_1), \ldots, Y^F(t_n)) \) be a trajectory of the TF (in molecule numbers) derived using an appropriate stochastic simulation algorithm (e.g. pcLNA). Define the continuous-time function \( Y^F(t), t \geq 0 \) of the observed at discrete times \( Y^F \) using an appropriate (smooth) interpolation method. Also define the degradation, \( u_{degr}(\delta, Y^g) \), and transcription, \( u_{tran}(\tau, Y^F) \), rate functions respectively. The algorithm proceeds as follows:

1. initialise the gene expression \( Y^g := Y^g(0) \) and set \( t := 0 \),
2. generate the next time of degradation event, \( t_1 \sim t + \text{Exp}(u_{degr}(\delta, Y^g)) \),
3. generate \( u \sim U([0,1]) \) and integrate the transcription rate function up to last time \( t_2 \) for which \( I = \int_{t}^{t_2} u_{tran}(\tau, Y^F(s))ds + \log(u) \leq 0 \).
4. Set \( t := (t_1 \wedge t_2) \) and
   \[
   Y^g := \begin{cases} 
   Y^g - 1, & \text{if } t = t_1, \\
   Y^g + 1, & \text{otherwise}
   \end{cases}
   \]
5. If $t < T_{\text{max}}$ return to step 2.

Note that a Gibson-Bruck type of this algorithm (see [17] [18]) could be used alternatively, but will make little difference for the purposes of our demonstration.

In our simulation, the degradation reaction rate of EGR-1 and COX-2 are

$$u_{\text{degr},g_1} = \delta_1 Y^{(g_1)}, \quad u_{\text{degr},g_2} = (\Omega \times \delta_2) \frac{Y^{(g_2)}}{Y^{(g_2)} + k_{g_2}}$$

and the transcription rates

$$u_{tr} = (\Omega \times \tau) \frac{Y_A^h}{Y_A^h + k_A} \frac{k_R^h}{Y_R^b + k_R^b}$$

where $Y_A$ the state of the activator TF and $Y_R$ the state of the inhibitor TF. As shown in Fig 4(a), the activator of EGR1 is the (unmodified) nuclear NF-κB , $N_n$, and the inhibitor its modified form, $(N_m)_n$ and for COX-2 vice versa. The parameter values used for EGR1 are $\delta_1 = 1 \times 10^{-3}$, $\tau = 8 \times 10^{-7}$, $k_A = 0.03$, $k_R = 0.03$ and for COX-2 are $\delta_2 = 5 \times 10^{-6}$, $\tau = 2 \times 10^{-7}$, $k_A = 0.015$, $k_R = 0.005$, $k_{g_2} = 1 \times 10^{-3}$, also $h = 5$.

4.7 Fisher Information Matrix (FIM) under the pcLNA

To compute FIM under the pcLNA, we need to compute derivatives of the solution $x = g(t)$, and the matrices $C(t_0, t)$, $V(t_0, t)$ with respect to the parameter vector $\theta$.

4.7.1 Derivatives for FIM computation

Consider a solution $x = \eta(t; t^*, x^*, \theta)$, $t \geq 0$, of the ode in ([19]) with initial condition $\eta(t^*; t^*, \theta) = x^*$. Here we write the solution as $\eta(t; t^*, x^*, \theta)$ (and the tangent vector as $F(\eta(t; \theta))$) to emphasise its dependence to the initial conditions $t^*$, $x^* = (x^*_1, x^*_2, \ldots, x^*_n)^T$ and the parameter vector $\theta = (\theta_1, \theta_2, \ldots, \theta_k)^T$.

The Jacobian matrix $J(x) = J(x; \theta)$ at $x = \eta(t; t^*, x^*, \theta)$ has entries

$$J_{i'i'} = \left(\frac{\partial F_i}{\partial x'}\right)_{i,i'} = \frac{\partial F_i}{\partial x'}.$$ 

Define the first and second derivatives of the solution $x = \eta(t; t^*, x^*, \theta)$ with respect to the initial conditions and parameters

$$u_k = \frac{\partial \eta}{\partial x_k^*}, \quad v_j = \frac{\partial \eta}{\partial x_j^*}, \quad w_{jk} = \frac{\partial u_j}{\partial x_k^*} = \frac{\partial u_k}{\partial x_j^*} = \frac{\partial^2 \eta}{\partial x_k^* \partial x_j^*}. \quad (31)$$

As we demonstrate below, the first derivative provides the columns of the fundamental matrix $C(t^*, t)$. The second provides the derivatives of the ode solution wrt to the parameters. The third provides the derivatives of the entries of the fundamental matrix wrt to the $j$-th parameter. All three are useful in computing the derivatives required for the FIM computation.

The derivative $u_k$ in (31) is the solution of the initial value problem

$$x' = J(t)x, \quad x(t^*) = e_k,$$

where $e_l = 0$, for $l \neq k$, $e_k = 1$, for $l = k$. 

29
Proof.

\[
\frac{d\eta}{dt} = F(\eta, \theta)
\]
\[
\frac{d}{d\mathbf{x}_k^*} \left( \frac{d\eta}{dt} \right) = \frac{dF}{d\mathbf{x}_k^*}(\eta, \theta)
\]
\[
\frac{d}{dt} \left( \frac{\partial \eta}{\partial \mathbf{x}_k^*} \right) = \frac{\partial F}{\partial \eta} \frac{\partial \eta}{\partial \mathbf{x}_k^*}
\]
\[
\frac{d\mathbf{u}_k}{dt} = J(t)\mathbf{u}_k
\]

and

\[
\mathbf{u}_k(t^*) = \frac{\partial \eta}{\partial \mathbf{x}_k^*}(t^*) = \frac{\partial \mathbf{x}_*}{\partial \mathbf{x}_k^*} = e_k. \Box
\]

Note that the principal fundamental matrix \( C \) is the solution of the latter initial value problem. The matrix can be written as

\[
C(t^*, t) = (u_1(t), u_2(t), \ldots, u_n(t)).
\]

Furthermore, \( v_j \) in (31) is the solution of the initial value problem

\[
x' = J(t)x + h_j(t), \quad x(t^*) = 0
\]

where \( h_j(t) = \frac{\partial F}{\partial \theta_j} \) (note: this is the partial derivative of \( F \) to \( \theta \), i.e. \( x \) is taken as constant to \( \theta \)).

Proof.

\[
\frac{d\eta}{dt} = F(\eta, \theta)
\]
\[
\frac{d}{d\theta_j} \left( \frac{d\eta}{dt} \right) = \frac{dF}{d\theta_j}
\]
\[
\frac{d}{dt} \left( \frac{\partial \eta}{\partial \theta_j} \right) = \frac{\partial F}{\partial \eta} \frac{\partial \eta}{\partial \theta_j} + \frac{\partial F}{\partial \theta_j} \frac{\partial \theta_j}{\partial \theta_j}
\]
\[
\frac{d\mathbf{v}_j}{dt} = J\mathbf{v}_j + h_j(t).
\]

Also,

\[
\mathbf{v}_j(t^*) = \frac{\partial \eta}{\partial \theta_j}(t^*) = \frac{\partial \mathbf{x}_*}{\partial \theta_j} = 0 \Box
\]

Furthermore, \( w_{jk} \) in (31) can be derived as the solution of the initial value problem,

\[
\frac{dx}{dt} = Jx + \left( \frac{\partial J_{kk'}}{\partial \eta} v_j + \frac{\partial J_{kk'}}{\partial \theta_j} \right)_{k,k'} \mathbf{u}_k, \quad x(t^*) = 0
\]

Proof.

\[
\frac{d}{d\theta_j} \frac{d\mathbf{u}_k}{dt} = \frac{d}{d\theta_j} (J(\eta, \theta)\mathbf{u}_k)
\]
\[
\frac{d}{dt} \frac{d\mathbf{u}_k}{d\theta_j} = \frac{dJ(\eta, \theta)}{d\theta_j} \mathbf{u}_k + J(\eta, \theta) \frac{d\mathbf{u}_k}{d\theta_j}
\]
\[
\frac{d\mathbf{w}_{jk}}{dt} = \frac{dJ(\eta, \theta)}{d\theta_j} \mathbf{u}_k + J(\eta, \theta)\mathbf{w}_{jk}
\]
where
\[
\frac{dJ(\eta, \theta)}{d\theta_j} = \left( \frac{dJ_{kk'}(\eta, \theta)}{d\theta_j} \right)_{k,k'}
\]
and
\[
\frac{dJ_{kk'}(\eta, \theta)}{d\theta_j} = \frac{\partial J_{kk'}}{\partial \eta_j} + \frac{\partial J_{kk'}}{\partial \theta_j} v_j.
\] (34)

Also,
\[
w_{jk}(t^*) = \frac{d}{d\theta_j} u_k(t^*) = \frac{d}{d\theta_j} e_k = 0.
\]

Finally, if \( V = V(t^*, t) \) in (25), the \( n \times n \) matrix \( X_j = dV/d\theta_j \) can be derived as the solution of
\[
\frac{dX_j}{dt} = X_jJ^T + JX_j + V \left( \frac{\partial J_{kk'}}{\partial \eta_j} v_j + \frac{\partial J_{kk'}}{\partial \theta_j} \right)_{k,k'} + \left( \frac{\partial J_{kk'}}{\partial \eta_j} v_j + \frac{\partial J_{kk'}}{\partial \theta_j} \right)_{k,k'} V
+ M \text{ Diag} \left( \frac{\partial u_l^T}{\partial \eta_j} v_j + \frac{\partial u_l^T}{\partial \theta_j} \right)_l M^T, \quad X_j(t^*) = 0
\] (35)

**Proof.** We differentiate (25) wrt to \( \theta_j \) to get
\[
\frac{d}{d\theta_j} \left( \frac{dV}{dt} \right) = \frac{d}{d\theta_j} \left( VJ^T + JV + MUM^T \right)
\]
\[
\frac{dX_j}{dt} = X_jJ^T + V \frac{dJ}{d\theta_j} + \frac{dJ}{d\theta_j} V + J(t)X_j + M \frac{dU}{d\theta_j} M^T
\]
The derivative \( dJ/d\theta_j \) is derived in (34), while \( \frac{\partial U}{\partial \eta_j} \) is a diagonal matrix with main diagonal entries,
\[
\frac{\partial u_i^T}{\partial \eta_j} v_j + \frac{\partial u_i^T}{\partial \theta_j} v_j + \frac{\partial r_l^T}{\partial \theta_j} \]

### 4.8 FIM computation for pcLNA joint distributions

The multivariate normality of the pcLNA joint distributions \( P_{\text{pcLNA}}(Q_{x_1}, Q_{x_2}, \ldots, Q_{x_k} | X(0) \sim MVN(\mu_0, \Sigma_0)) \), for \( k = 2, \ldots, \), allows us to compute the corresponding FIM as in (7) by computing the derivatives of its mean vector \( \mu \) and precision matrix \( A \) in (29) and (30), respectively.

As part of this, the derivatives of \( \partial x(t_i)/\partial \theta, \partial C(t_i, t_{i+1})/\partial \theta, \) and \( \partial V(t_i, t_{i+1})/\partial \theta \) are derived by solving the initial value problems in (32), (33) and (35), respectively. The derivatives \( \partial R_t^{(2)}/\partial \theta \) are derived by using the fsolve function of MATLAB to derive a solution of the linear problem with constraints defined by differentiating the equations \( F(x(t))^T R_t^{(2)} = 0 \) and \( R_t^{(2)T} R_t^{(2)} = I \).

### 4.9 Effects of observing less time-points on the singular values of the FIM

The singular values of the FIM in Fig. 13(b) are computed at time-points \( t = 12, 22, 110, 142, 198, 244, 286, 344 \) min. These times include the first four peak times of the nuclear NF-κB concentration, \( t \approx 22, 142, 244, 344 \) min, an earlier and three times between the peaks. If instead the singular values of the FIM are measured at time-points \( t = 12, 22 \) min, then the singular values are displayed in Fig. 11. The singular values are slightly increased compared to the base model, but the increase is clearly less prominent, particularly for the second singular value, compared to the corresponding singular values for the times \( t = 12, 22, 110, 142, 198, 244, 286, 344 \) min.
Figure H: The singular values of the NF-κB base model (black -×- line) and the mNF-κB model (red -o- line) for observations taken at times \( t = 12, 22 \) min.

5 NF-κB target genes

We used the list of NF-κB target genes provided in [5]. This comprises of the list of target genes taken from the Boston University web site at http://www.bu.edu/nf-kb/gene-resources/target-genes/ combined with target genes found via an extensive ChIP-seq search performed in [15]. There are a total of 1361 genes in this list.

The profile data of the 8 target genes are displayed in Fig. I and J. Fig. K illustrates how these 8 genes can be used to distinguish between 14 different experimental conditions. An approximate expression profile alphabet is provided, which includes profiles that can be described as no response, or early/medium/late, sustained or not sustained response. These profiles are identified by clustering [19]. Then we see on the table that each of these 14 different experimental conditions gave a different response combination over these 8 genes.

---

[1] accessed 8th August 2015
Figure I: Microarray profiles of the target genes presented in Fig 2. The condition is given in the legend. Some genes have multiple microarray data.
Figure J: Microarray profiles of the target genes presented in Fig 2. The condition is given in the legend. Some genes have multiple microarray data.
**Figure K:** Summary table of the expression of 8 genes (columns) that can distinguish the 14 different experimental conditions (rows). Each letter corresponds to an approximate expression profile given on the right. This includes profiles that can be described as either no response (b), early, medium or late, transient or sustained response. Reading the expression table from left to right and top to bottom, the colours emphasise the gene that identifies each condition. The profile data are given in Fig I and J.

**References**

[1] Anderson DF, Kurtz TG. Continuous Time Markov Chain Models for Chemical Reaction Networks. In: Koeppl H, Setti G, di Bernardo M, Densmore D, editors. Design and Analysis of Biomolecular Circuits: Engineering Approaches to Systems and Synthetic Biology. New York: Springer; 2011.

[2] Ashall L, Horton CA, Nelson DE, Paszek P, Harper CV, Sillitoe K, et al. Pulsatile stimulation determines timing and specificity of NF-κB-dependent, transcription. Science 2009; 324(5924):242–6.

[3] Cover TM, Thomas JA. Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing). Wiley-Interscience New York, NY, USA; 2006.

[4] Gillespie DT. Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry. 1977 Dec; 81(25):2340–2361.

[5] Harper CV, Woodcock DJ, Lam C, Garcia-Albornoz M, Adamson A, Ashall L et al. Temperature regulates NF-κB dynamics and function through timing of A20 transcription. Proc Natl Acad Sci USA. 2018; 115(22): E5243-E5249.

[6] Hartman P. Ordinary differential equations. New York: John Wiley & Sons, Inc. London and Sydney: Wiley; 1964.

[7] Horn, RA, Johnson, CR. Topics in Matrix Analysis. Cambridge University Press; 1991.

[8] Kurtz TG. Approximation of Population Processes. Regional Conference Series Applied Mathematics vol36. Society for Industrial and Applied Mathematics, Philadelphia, PA; 1981.

[9] Kurtz TG. Limit Theorems for Sequences of Jump Markov Processes Approximating Ordinary Differential Processes. Journal of Applied Probability. 1971 Jun;8(2), 344–356.

[10] van Kampen NG. Stochastic Processes in Physics and Chemistry, Third Edition. Amsterdam: Elsevier. Boston and London: Elsevier; 2007.
[11] MATLAB 7.11 (R2017b), The MathWorks, Inc, Natick, Massachusetts, United States.

[12] Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, et al. TRANSFAC
and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids
Res. 2006 Jan 1;34(Database issue):D108-10. doi: 10.1093/nar/gkj143.

[13] Minas, G, Rand, DA. Long-time analytic approximation of large stochastic oscillators:
Simulation, analysis and inference. PLOS Computational Biology. 2017; 13(7), e1005676.

[14] Numerical Recipes: The Art of Scientific Computing. Third Edition, 1256 pp. Cambridge
University Press; 2007.

[15] Satoh J. Molecular network of ChIP-Seq-based NF-κB p65 target genes involves diverse
immune functions relevant to the immunopathogenesis of multiple sclerosis. Mult. Scler.
Relat. Disord. 2014; 3(1): 94–106.

[16] Tay, S, Hughey, JJ, Lee, TK, Lipniacki, T, Quake, SR, Covert, MW. Single-cell NF-κB
dynamics reveal digital activation and analogue information processing. Nature. 2010;
466(7303): 267–271.

[17] Wilkinson DJ. Stochastic modelling for systems biology. Boca Raton: CRC Press/Taylor
& Francis; 2012.

[18] Gibson MA, Bruck J. Efficient Exact Stochastic Simulation of Chemical Systems with
Many Species and Many Channels The Journal of Physical Chemistry A. 2000; 104(9):
1876–1889.

[19] Neal, RM. Markov Chain Sampling Methods for Dirichlet Process Mixture Models.
Journal of Computational and Graphical Statistics. 2000; 9(2), 249.