Accurate information transmission through dynamic biochemical signaling networks

Jangir Selimkhanov, Brooks Taylor, Jason Yao, Anna Pilko, John Albeck, Alexander Hoffmann, Lev Tsimring, Roy Wollman

Motivation

- Cell signaling lends itself very naturally to information theory analysis
- Response to paper published in recently:
  - R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, A. Levchenko, Information transduction capacity of noisy biochemical signaling networks. Science 334, 354–358 (2011).
  - Analyzed information transmission in TNF pathway
  - Showed that information between extracellular signal and cellular response is less that 1 bit.
  - Cellular response only measured with scalar variables

Introduction

Compare the Information Transmission in Cellular Responses

- Signal to Noise Ratio (SNR)
- Mutual Information

- Scalar Responses
- Dynamic Responses

For Signaling Networks

- ERK
- Ca^{2+}
- NF-κB

With Stochasticity:

- Intrinsic Noise
- Extrinsic Noise

Overview

- Experimental Cellular Response Measurement
- Information Transmission Analysis
- Model Simulation and Analysis of Noisy Signaling (Erk Only)

	Erk	Ca^{2+}	NFκB
Measure	EKARev-NES FRET reporter	Fluo-4 indicator dye	translocation of EYFP-p65
Stimulus	EGF	ATP	LPS
Doses (#)	8-16 / expmt	6 / expmt	9 / expmt
Cells	825,001	80,566	4,554
Time (step)	60 min (1-3 min)	15 min (3 sec)	18 hr (5 min)

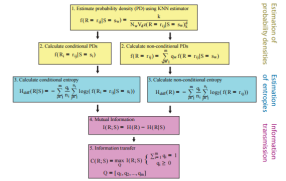
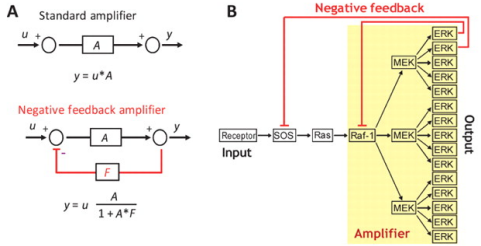


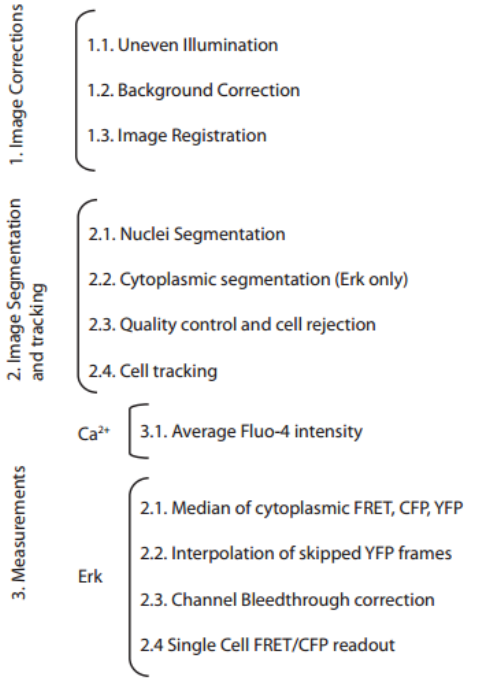
Fig. S7
General scheme for estimation of information transmission based on experimentally obtained conditional responses (R) to scalar input levels (S).



Experimental Cellular Response Measurement

	Erk	Ca^{2+}	NFκB
Measure	EKARev-NES FRET reporter	Fluo-4 indicator dye	translocation of EYFP-p65
Stimulus	EGF	ATP	LPS
Doses (#)	8-16 / expmt	6 / expmt	9 / expmt
Cells	825,001	80,566	4,554
Time (step)	60 min (1-3 min)	15 min (3 sec)	18 hr (5 min)

Image Analysis (ERK and Ca^{2+})



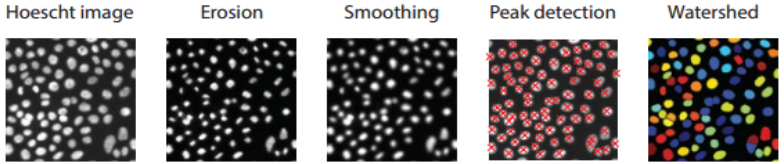


Fig. S2
Image analysis steps done to segment nuclei using Hoescht staining. Only a small part of an image is shown for brevity.

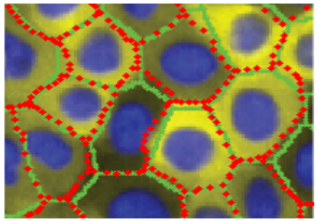


Fig. S3
MCF10a cells stained with Hoescht (blue) and expressing EKARev (yellow) with boundaries between cells determined by automatic segmentation marked in green. Red dots indicate "true" boundaries identified by eye. The error in cell segmentation based automatically obtained and "true" boundaries is $9\% \pm 2\%$ (SEM).

Image Analysis (NF-κB)

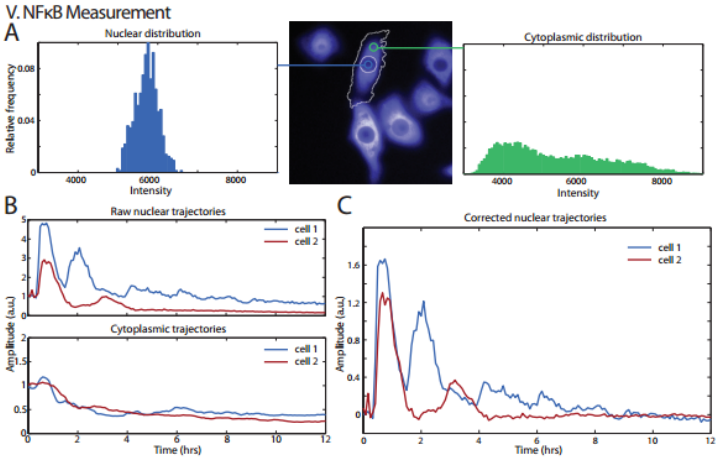


Fig. S6
NFκB measurement. (A) We calculate values that correspond to the nuclear (blue, median/mode) and cytoplasmic (green, higher mode) intensity distributions, which show identical decreasing trends over time (as a function of changing cell morphology). (B) Raw nuclear trajectories show cells that are at basal level before stimulation, and eventually stabilize after a maximum 10-14 hrs. We use this information, along with the shape computed from each cell's cytoplasmic trajectory, to calculate a true baseline for each cell. (C) Final, corrected and normalized nuclear trajectories can be directly compared.

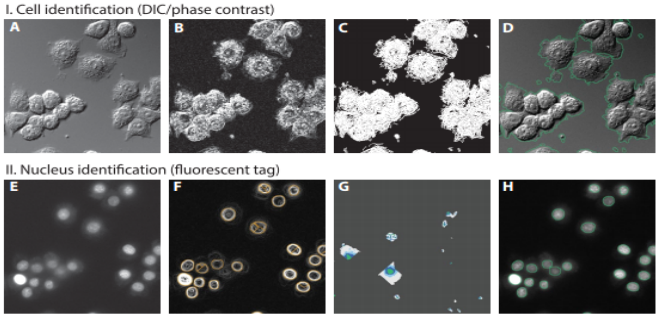


Fig. S4
Initial cell and nucleus identification. (A) Raw DIC image. (B) Sobel edge-magnitude image. (C) Thresholded edge image. (D) Final foreground/cellular boundaries. (E) Raw nuclear image (H2B-mCherry). (F) Strong objects found by scanning edge image (G) Weaker objects found in remaining candidate areas - pixels are ranked by intensity, and appropriately concentric objects are identified as nuclei. (H) Final nuclear boundaries.

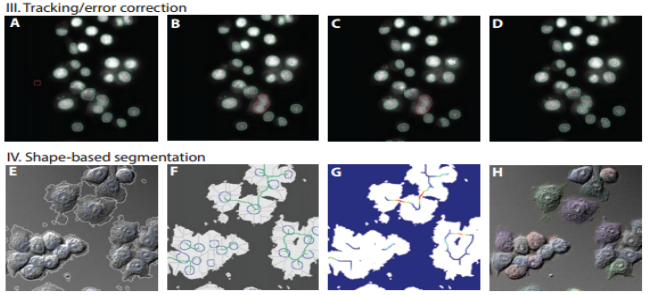
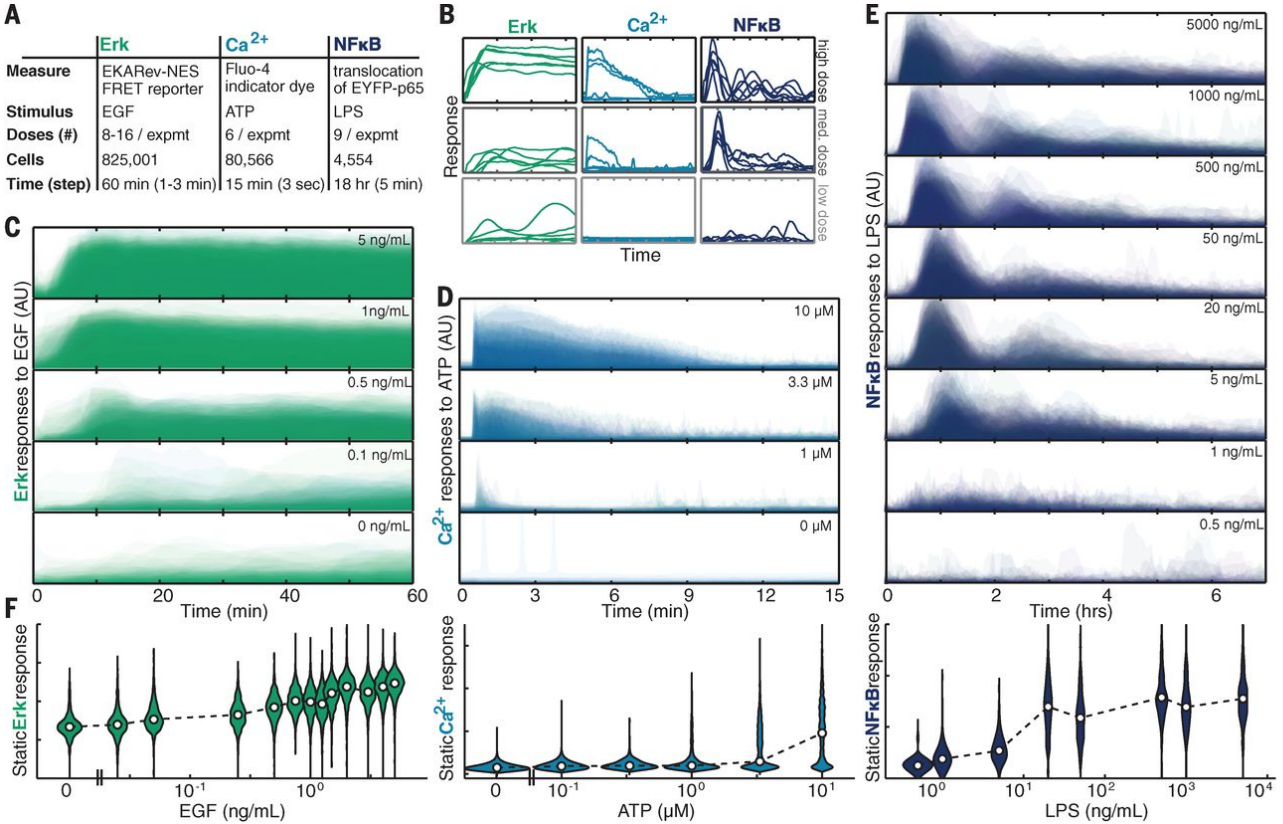


Fig. S5
Tracking and segmentation. (A-D) 4 consecutive images are individually processed, then tracked together. Voting on objects across the stack allows easy identification and correction of false positives (A, red) and false negatives (B and C, red). (E) Segmentation begins using the nuclear and cellular boundaries identified earlier. (F) The morphological skeleton is computed, then pruned to areas connecting each nuclei. (G) We identify the local maxima of the distance transformation along the pruned skeleton, as candidate splitting points. (H) Final segmented image.

Image Analysis Results



Mutual Information

General

- Measurement of mutual dependence of two variables
- Has a negative correlation to the similarity between a joint distribution of 2 variables and the product of their marginal distributions.
- When calculated using log base 2 has the unit of bits.

In paper

- Measurement of the mutual dependence of the extracellular ligand signal (S) and the cellular response (R)

$$S = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_i \\ s_m \end{bmatrix}, R = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_i \\ \vdots \\ R_m \end{bmatrix}, s_i \rightarrow R_i = \begin{bmatrix} r_{i1} \\ r_{i2} \\ \vdots \\ r_{ij} \\ \vdots \\ r_{in_i} \end{bmatrix}, r_{ij} = [r_{ij,1}, r_{ij,2}, \ldots, r_{ij,d}]$$

- m = # of levels of the extracellular ligand signal concentration
- n = # of timepoint measurements

Information Transmission Calculation

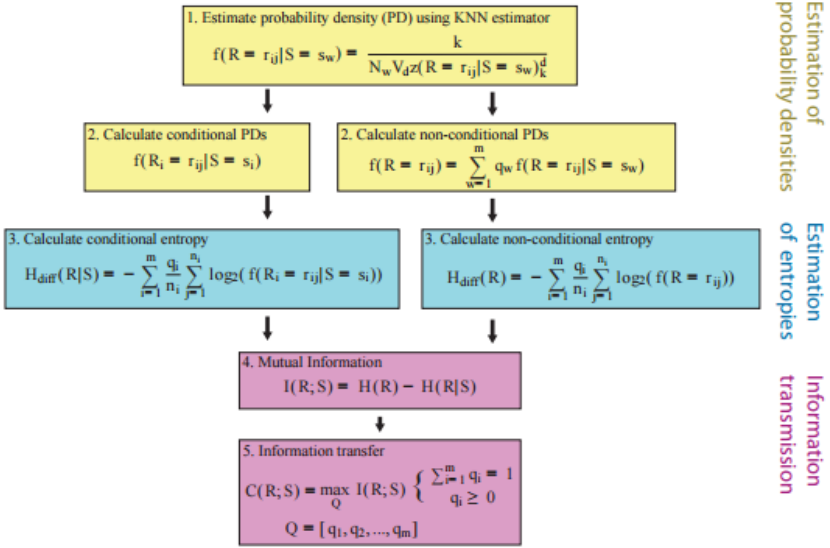


Fig. S7
General scheme for estimation of information transmission based on experimentally obtained conditional responses (R) to scalar input levels (S).

How Mutual Information was Calculated

$$I(R;S) = H(R) - H(R|S).$$

$$H_{\text{diff}}(X) = -\int_{-\infty}^{\infty} f(x) log_2(f(x)) dx.$$

$$H_{\text{diff}}(X) = -\int_{0}^{1} log_2(f(x)) dy. \tag{2.3}$$

where $y = \int_{-\infty}^{x} f(t))dt$ is the cumulative probability density. We can estimate y by the cumulative probability distribution of N_x observations using

$$H_{\text{diff}}(X) = -\sum_{j=1}^{N_x} \delta_j log_2(f(x_j)), \tag{2.4}$$

where δ_j is the probability of observing x_j, $P(X = x_j)$.

How Mutual Information was Calculated

$$I(R;S) = H(R) - H(R|S).$$

$$f(R = r) = \sum_{w=1}^{m} q_w f(R = r|S = s_w).$$

$$H_{\text{diff}}(R_i|S = s_i) = -\sum_{j=1}^{n_i} \frac{1}{n_i} log_2(f(R_i = r_{ij}|S = s_i)).$$

$$H_{\text{diff}}(R) = -\sum_{i=1}^{m}\sum_{j=1}^{n_i} \delta_{ij} log_2(f(R = r_{ij})).$$

$$H_{\text{diff}}(R|S) = \sum_{i=1}^{m} q_i H_{\text{diff}}(R_i|S = s_i) = -\sum_{i=1}^{m} q_i \sum_{i=1}^{n_i} \frac{1}{n_i} log_2(f(R_i = r_{ij}|S = s_i)).$$

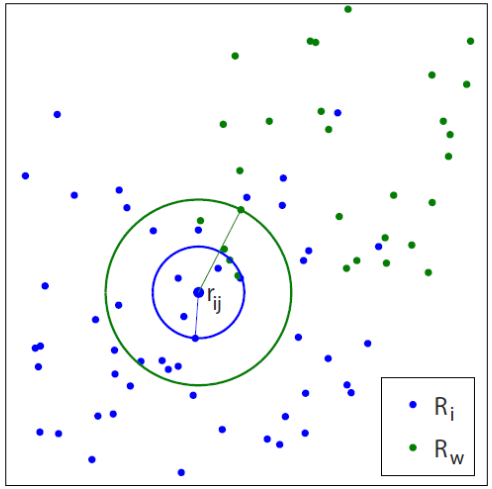
$$H_{\text{diff}}(R) = -\sum_{i=1}^{m} \frac{q_i}{n_i} \sum_{j=1}^{n_i} log_2(f(R = r_{ij})).$$

4. Mutual Information

$$I(R;S) = H(R) - H(R|S)$$

q = probability of the ligand concentration

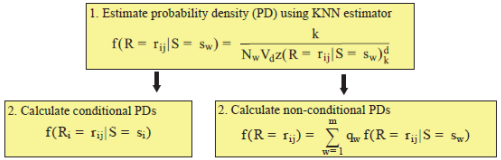
KNN Estimator



$$f(x_j|X) = \frac{k}{N_x V_d z(x_j|X)_k^d}$$

$$V_d = \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+1)}$$

Fig. S8
Representation of k-nearest neighbor calculation for $k = 5$. The blue circle radius is the distance to the fifth closest neighbor within the same input response represented by blue points. The green circle radius is the distance to the fifth closest neighbor to a different input response (green points).



Plugging In KNN Estimate

$$I(R;S) = H(R) - H(R|S).$$

$$H_{\text{diff}}(R|S) = -\sum_{i=1}^{m} \frac{q_i}{n_i} \sum_{j=1}^{n_i} log_2(\frac{k}{n_i V_d z(r_{ij}|R_i)_k^d}).$$

$$H_{\text{diff}}(R) = -\sum_{i=1}^{m} \frac{q_i}{n_i} \sum_{j=1}^{n_i} log_2(\sum_{w=1}^{m} q_w \frac{k}{n_w V_d z(r_{ij}|R_w)_k^d})$$

3. Calculate conditional entropy

$$H_{\text{diff}}(R|S) = -\sum_{i=1}^{m} \frac{q_i}{n_i} \sum_{j=1}^{n_i} \log_2(f(R_i = r_{ij}|S = s_i))$$

3. Calculate non-conditional entropy

$$H_{\text{diff}}(R) = -\sum_{i=1}^{m} \frac{q_i}{n_i} \sum_{j=1}^{n_i} \log_2(f(R = r_{ij}))$$

Problem : We don't know q

We can use maximal information transfer instead.

$$C(R;S) = \max_{Q}\{I(R;S)\},$$

where $Q = [q_1, q_2, ..., q_m]$, such that $\sum_{i=1}^{m} q_i = 1$ and $q_i \geq 0$

5. Information transfer

$$C(R;S) = \max_{Q} I(R;S) \begin{cases} \sum_{i=1}^{m} q_i = 1 \\ q_i \geq 0 \end{cases}$$

$$Q = [q_1, q_2, ..., q_m]$$

Information Transmission Calculation Review

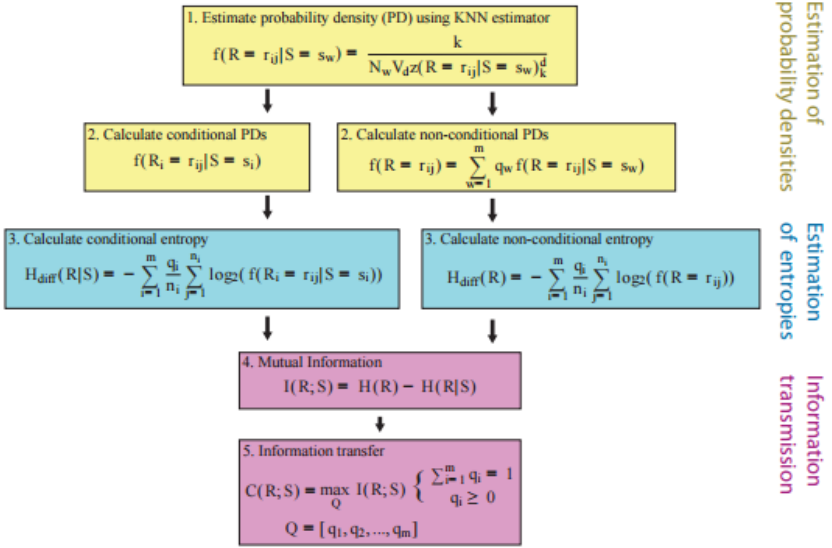
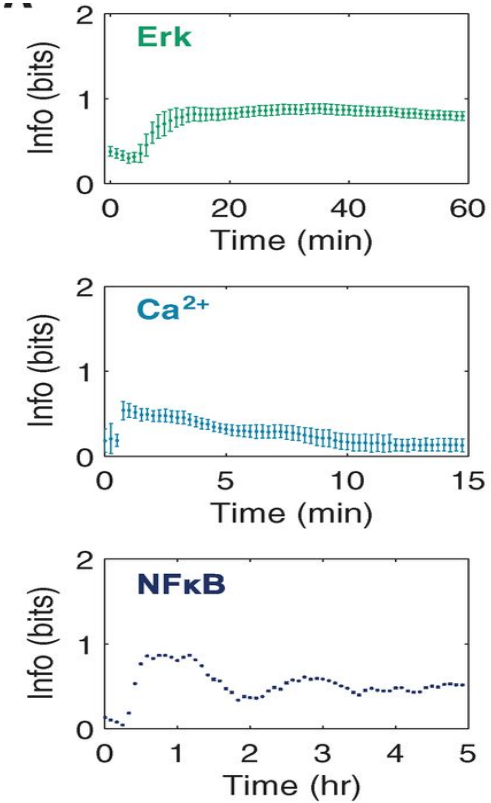
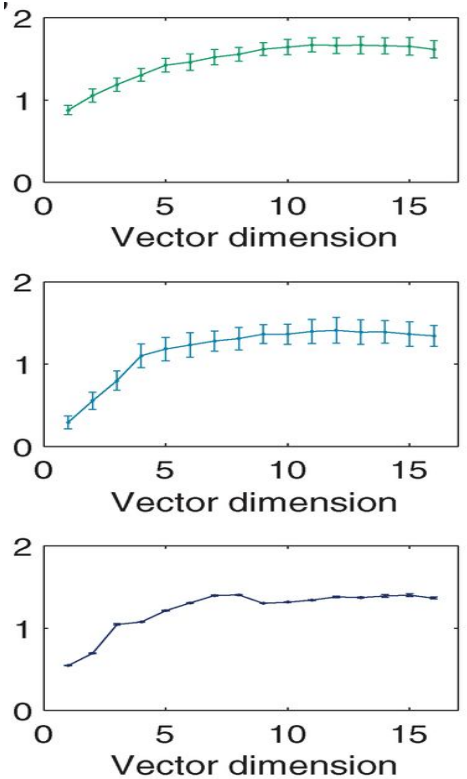


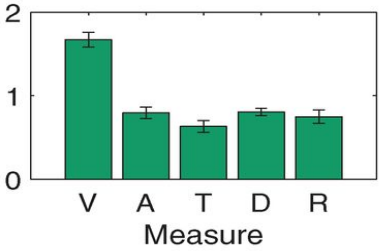
Fig. S7
General scheme for estimation of information transmission based on experimentally obtained conditional responses (R) to scalar input levels (S).

Single Time Point Mutual Information vs Dynamic Mutual Information

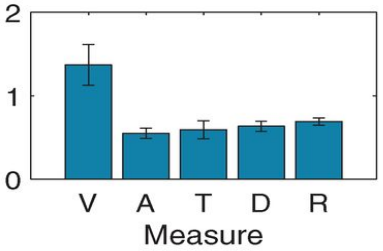




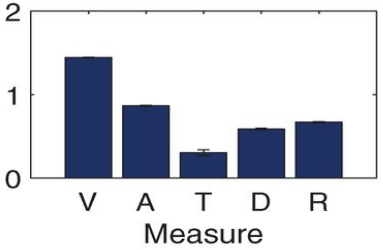
Scalar Mutual Information vs Dynamic Mutual Information



ERK



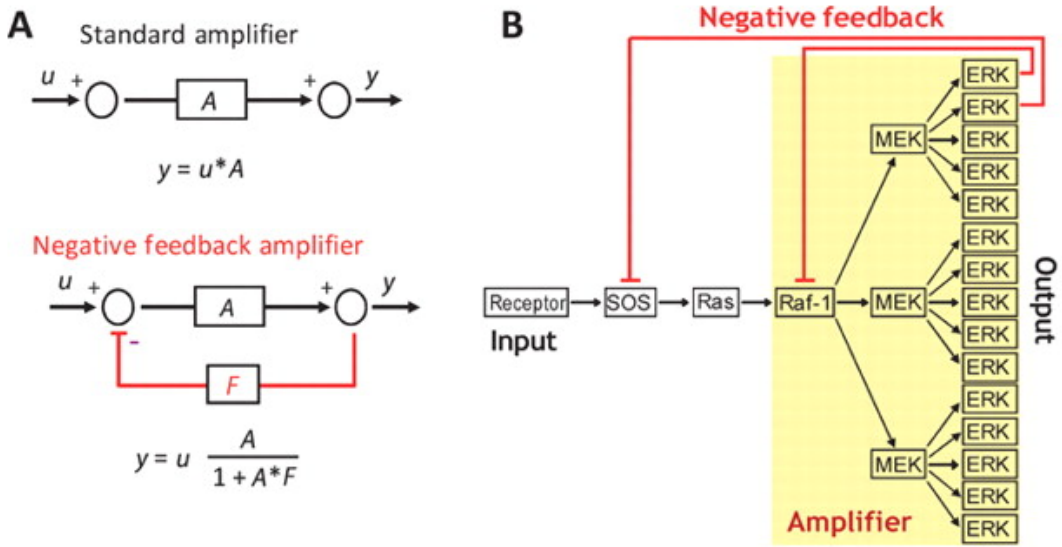
Ca^{2+}



NF-κB

Comparison of the multivariate vector (V) measurement to the following scalar responses: maximum response amplitude (A), maximum response time (T), maximal rate of response (D), ratio of maximum response amplitude to initial response amplitude (R). Error bars are SEMs from six biological replicates for ERK and four for Ca^{2+}, and SDs from five jackknife iterations for NF-κB

Simulation of Noisy Signaling



O. E. Sturm, R. Orton, J. Grindlay, M. Birtwistle, V. Vyshemirsky, D. Gilbert, M. Calder, A. Pitt, B. Kholodenko, W. Kolch, The mammalian MAPK/ERK pathway exhibits properties of a negative feedback amplifier. Sci. Signal. 3, ra90 (2010).

Does the model agree with the experimental results?

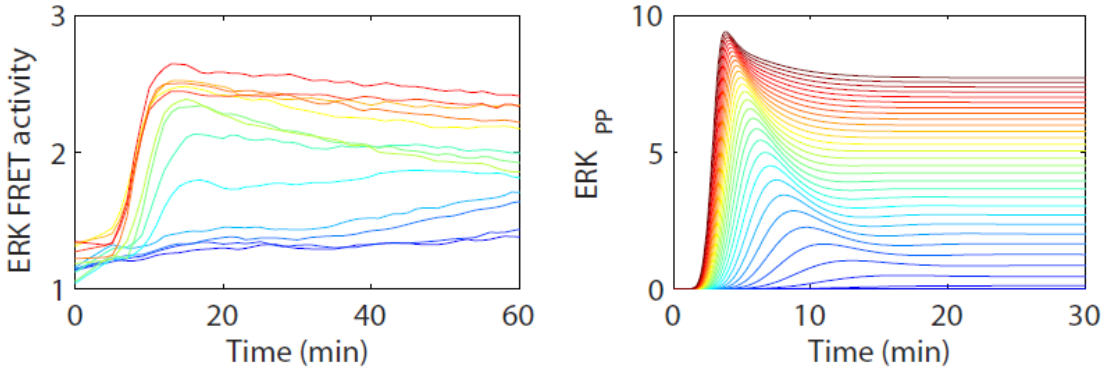


Fig. S15
Model simulation comparison to experimental ERK FRET trajectories. (A) Mean response of ERK FRET sensor to persistent EGF input. (B) ERKpp response trajectories from simulations of the ERK model (Sturm *et al*) for increasing amounts of RasGTP.

Adding Noise to the Model

Intrinsic Noise

- Stochasticity inherent to biochemical reactions
- Adds uncertainty in all time dimensions independently of one another
- Simulated in the model as a Gaussian random variable added to the response

Extrinsic Noise

- Variability in cellular states
- Fluctuations due to extrinsic noise in each time point are deterministically dependent on one another
- Simulated in the model by randomly selecting MEK and ERK values from a uniform distribution that varies a max of 20% from the values used by Sturm et. al

Effect of noise in the model on mutual information

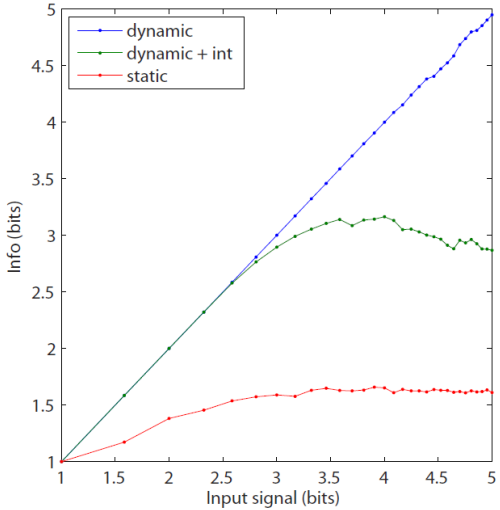
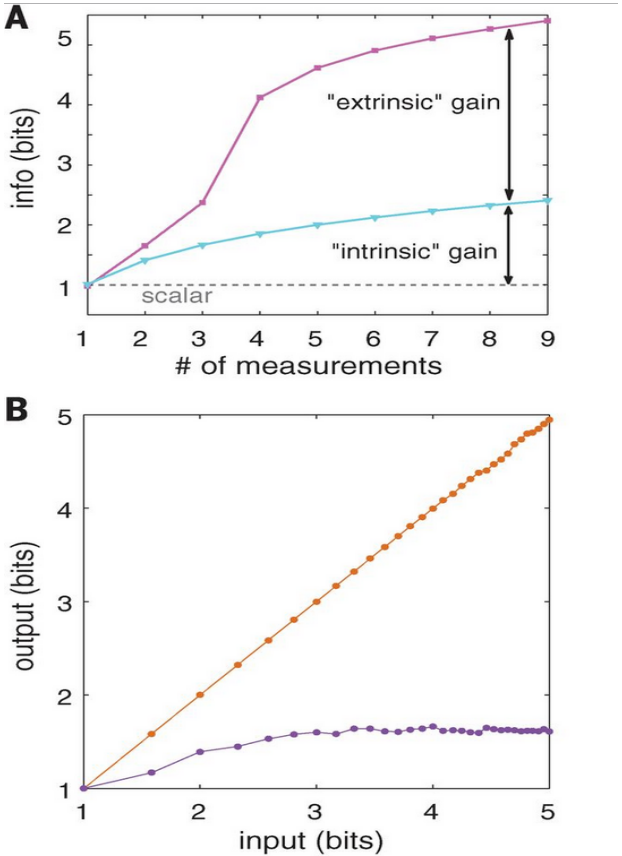


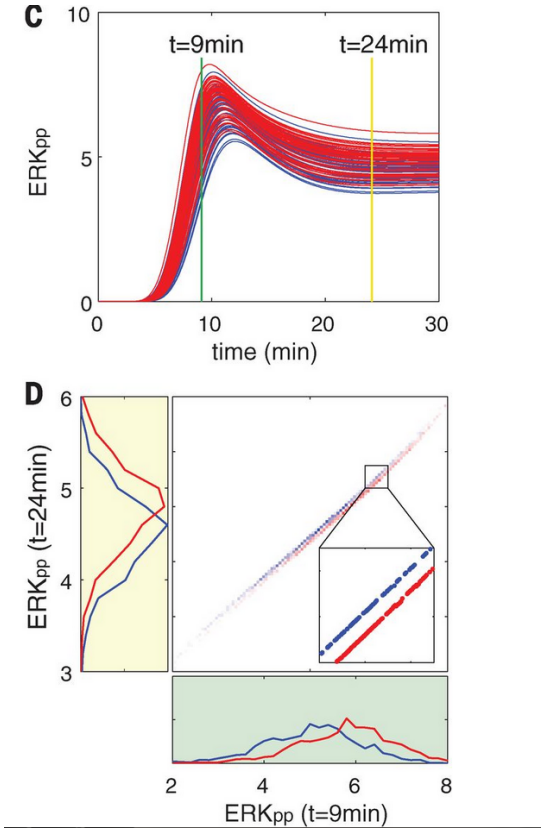
Fig. S16
Information transmission capacity of dynamic (blue, green) and static (red) calculated based on the full computational model of ERK where the extrinsic (all) noise and intrinsic (green) noise contributed to cell response variability.

Dynamic responses can eliminate information loss



- (A) Graphical representation of the analytical expression for the gain in mutual information from overcoming intrinsic (cyan) and extrinsic (magenta) noise sources obtained from random linear Gaussian inputs and outputs with three parameters .
- (B) Information transmission capacity of dynamic (orange) and static (maximal response, purple) responses calculated using simulated trajectories from the computational model of ERK with only the extrinsic noise contributing to cell response variability.

Dynamics vs Scalar Measurements (A closer look)



(C) Example of ERK trajectory variability for two different inputs levels (red and blue). Variability was generated using a uniform distribution of a single parameter, MEK values, that was varied by ±20%.

(D). Two-dimensional histogram (center) and marginal distributions (left and bottom) for the two input levels (shown in red and blue) at two time points (t = 9 and 24 min) from the trajectories in (C). Because only a single parameter was varied, the responses vary on a 1D curve. As a result, although the univariate marginal distributions show substantial response overlap, the 2D distribution shows completely separated response levels (inset).

Adding noise to an experimental model of ERK

- Intrinsic Noise is already is inherent to any real biological system

- Extrinsic Noise is created by adding an inhibitor of MEK (U0126) at different doses

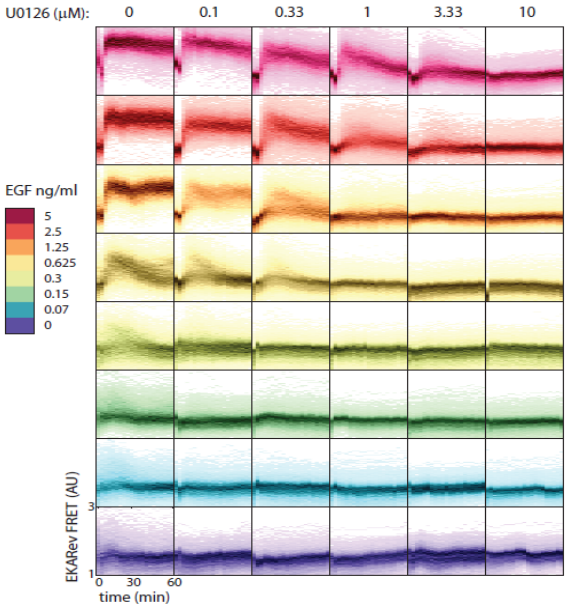


Fig. S23
Data from inhibitor experiments. The columns represent different MEK inhibitor (U0126) concentrations. The color-coded rows are different EGF induction levels. The plots show cell distribution with time where darker tint represents higher probability density. The addition of inhibitor leads to reduced ERK response due to decrease in signal propagation through the ERK pathway.

Estimating Noise from the Experimental Data

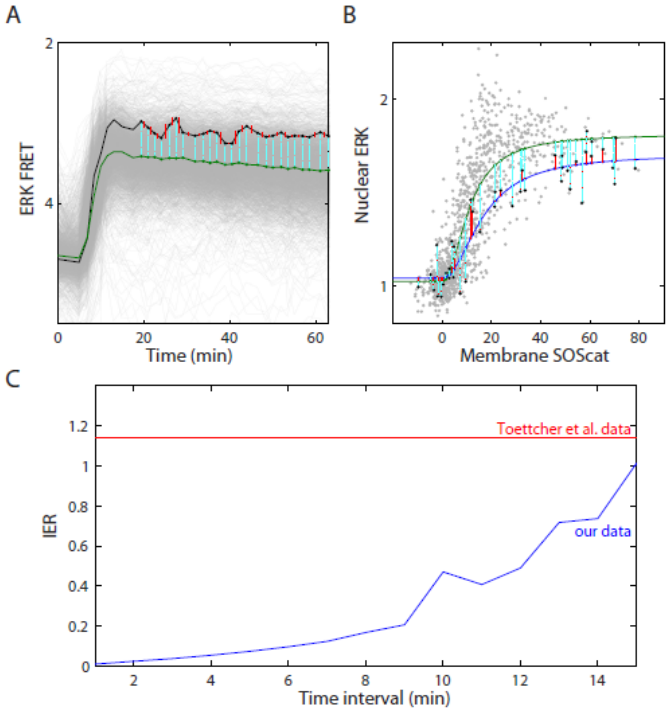


Fig. S21

(A) Using our data, intrinsic noise was estimated by the mean of the mean of squared errors between successive ERK trajectory points (red). Total noise was estimated by the mean of squared errors (cyan) between single ERK trajectory and average of all trajectories (green). Extrinsic ratio was obtained from the difference between total noise and intrinsic noise. The mean ratio of intrinsic to extrinsic noise was estimated to be 0:024.

(B)We fit a Hill function to the data and calculated the mean squared error between the fit for each cell (intrinsic noise) and between the fit for all points and each cell (total noise). The IER was estimated to be 1: 14

(C) For increasing the time step in our estimate we find an increase in our estimate of IER.

Signal to Noise Ratio(SNR)

SNR:

- Value that represent how strong the response is compared to how strong the noise is.
- We can use SNR as summary of the amount of noise in the signal.

4.3.1 Signal-to-Noise Ratio (SNR)

To calculate ERK signal-to-noise ratio (SNR), we defined the signal magnitude σ_r^2 as the variance of average responses over all m input levels of EGF:

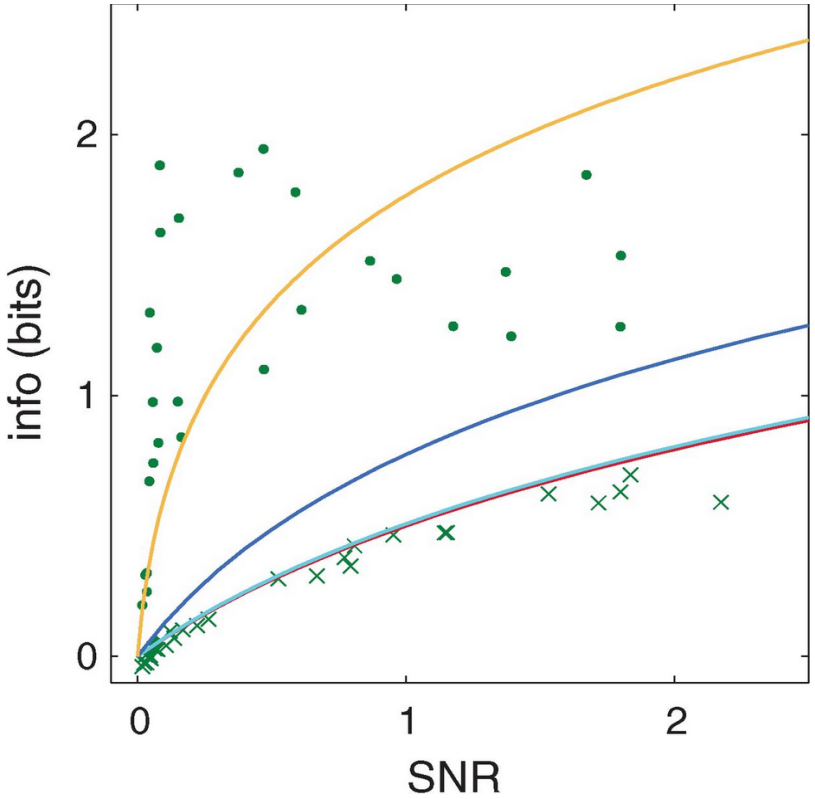
$$\sigma_r^2 = \frac{1}{m}\sum_{i=1}^{m}((\frac{1}{m}\sum_{w=1}^{m}\frac{1}{n_w}\sum_{j=1}^{n_w} r_{wj}) - \frac{1}{n_i}\sum_{j=1}^{n_i} r_{ij}))^2 \tag{4.1}$$

Noise magnitude was defined as the average of the variances of n_i responses to a single input level of EGF:

$$\sigma_n^2 = \frac{1}{m}\sum_{i=1}^{m}(\frac{1}{n_i}\sum_{j=1}^{n_i}(\frac{1}{n_i}\sum_{w=1}^{n_i} r_{iw} - r_{ij})^2 \tag{4.2}$$

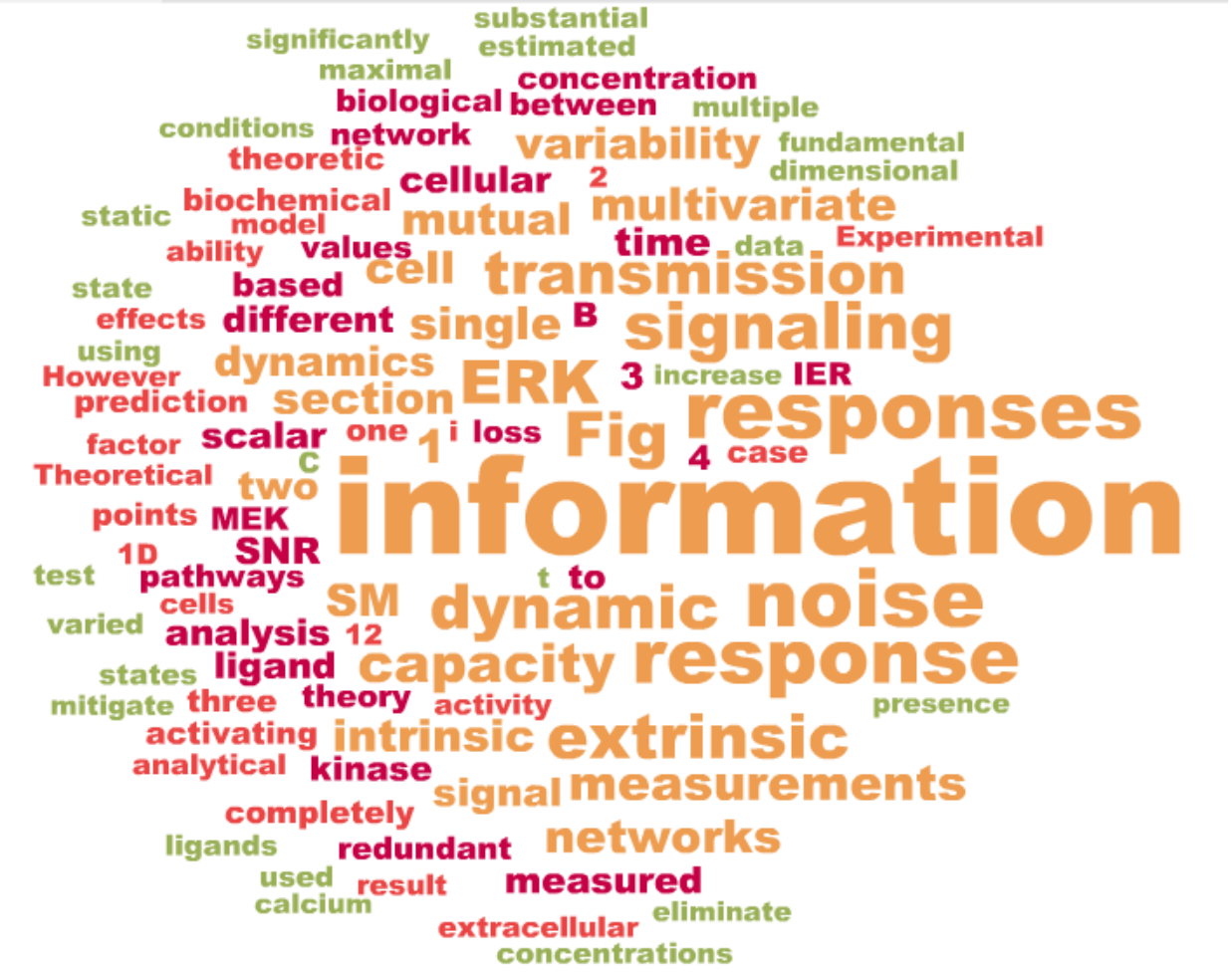
SNR is then σ_r^2/σ_n^2.

Comparing SNR to Information Transfer



Experimental measurement of the mutual information between ERK response and EGF measured as a function of the response signal-to-noise ratio (SNR). Each marker represents calculations of SNR and mutual information from the dynamic (dot) and maximal scalar (cross) responses of cells from an eight-well dose-response experiment. Data shown are calculated based on single-cell responses from 29 experiments with six doses of MEK inhibitor U0126 . Lines represent theoretical predictions of the mutual information as a function of SNR for three types of responses: static scalar (red line), redundant measurements where the multivariate response has no dynamics (dark and light blue lines) calculated based on two independent estimates of IER , and dynamic response (orange) that can mitigate both intrinsic and extrinsic noise.

Questions?


substantial
significantly
estimated
maximal
concentration
biological
between
multiple
conditions
network
variability
fundamental
theoretic
dimensional
cellular
2
biochemical
multivariate
static
model
mutual
time
data
Experimental
ability
values
cell
transmission
state
based
effects
different
single
B
signaling
using
dynamics
However
ERK
3
increase
IER
prediction
section
responses
factor
scalar
one
1
i
loss
Fig
4
case
C
Theoretical
two
information
points
MEK
1D
SNR
test
pathways
t
to
cells
SM
dynamic
noise
varied
analysis
12
capacity
response
states
ligand
mitigate
three
theory
activity
presence
activating
intrinsic
extrinsic
analytical
kinase
signal
measurements
completely
networks
ligands
redundant
used
result
measured
calcium
eliminate
extracellular
concentrations