optimal information transfer in a noisy nonlinear neuron

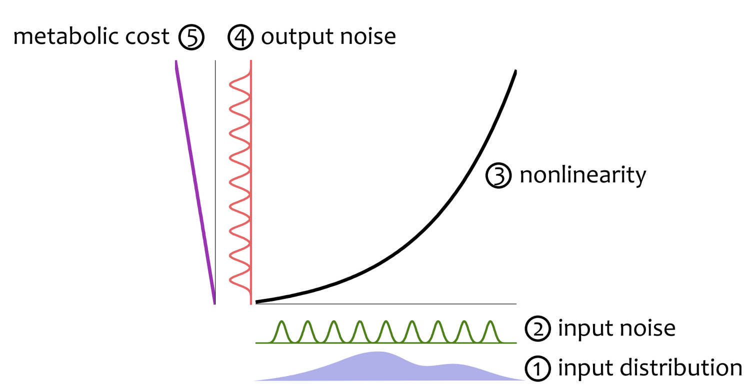
yan karklin and **eero p. simoncelli**
howard hughes medical institute and *new york university*

we consider a model for efficient coding in a single neuron that includes the primary attributes relevant to the biological problem: 1) non-Gaussian input statistics, 2) input noise, 3) non-linear neural processing, 4) output noise, and 5) a metabolic cost associated with spiking. *what nonlinear transfer function is best for information transmission under these conditions?*

to make the problem tractable, we make some assumptions about the smoothness of the nonlinearity and the input distribution. we then derive an analytic form for the optimal nonlinear transfer function, and apply this theory to data by testing the optimality of retinal ganglion cell (RGC) responses. the model yields good fits to RGC mean responses, as well as their variability. the resulting parameters suggest that differences between ON and OFF cell nonlinearities (Chichilnisky and Kalmar, 2002) are due to differences in both the pre-processing of their inputs and the metabolic costs of their responses.

a model noisy nonlinear neuron

the basic ingredients



1. **input distribution:** arbitrary smooth pdf
2. **input noise:** additive Gaussian*
3. **nonlinearity:** point-wise monotonic, smooth
4. **output noise:** additive Gaussian*
5. **metabolic cost:** proportional to output level

$$r = f(x + n_x) + n_r$$

(*) additive noise assumption can be relaxed to include Gaussian noise with input- or output-dependent variance

allow us to ask:

what nonlinearity maximizes the mutual information beween input and output?
what effect do noise and metabolic cost have on its shape?
what is the optimal output distribution?

maximize information conveyed about the stimulus, subject to a metabolic cost constraint on the output

$$L^* = \max E_x \left[I(x;r) - \lambda C(r) \right]$$

parameter λ: trades off information gained vs metabolic cost expended

computing mutual information

$$I(x;r) = H(x) - H(x|r)$$

$$L^* = \max E_x \left[-H(x|r) - \lambda C(r) \right]$$

how do we compute the conditional entropy?

we want to **minimize uncertainty** about the input for all response levels *r*, over the input distribution

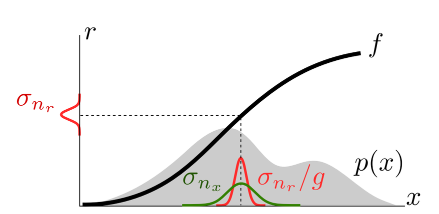
assume *f* is smooth and can be appoximated **locally** with a 1st order Taylor expansion, $f_i(x) \approx g_i x + c_i$

input-dependent "constant"

then, if input and output noise are additive Gaussian (and small w.r.t the curvature of *f*), we can compute the conditional entropy for each output r_i:

$$p(x|r_i) \propto p(r|x)p(x)$$

$$p(r|x) = \mathcal{N}(r; g_i x + c_i, g_i^2 \sigma_{n_x}^2 + \sigma_{n_r}^2)$$



for abitrary p(x), exact entropy of the posterior can be intractable

we replace the prior with Gaussian q(x) that matches the curvature at the mode of the posterior

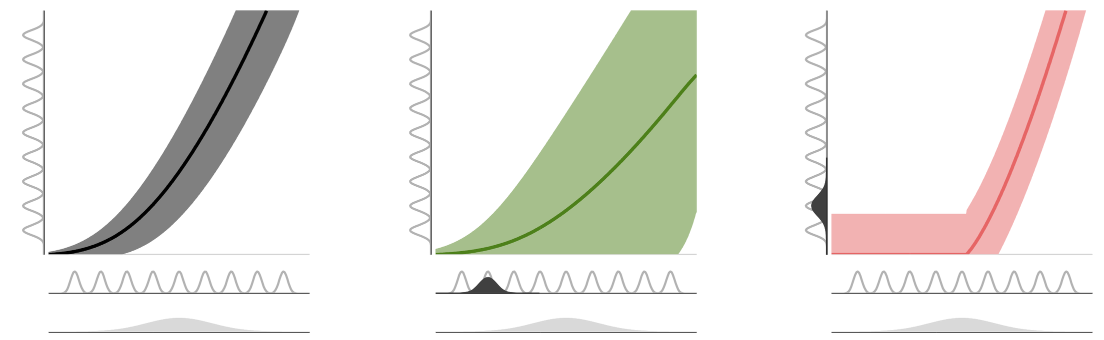
this yields a Gaussian with input-dependent variance

$$\sigma_{x|r}^2(x_i) = \left(\frac{1}{\sigma_{n_x}^2 + \sigma_{n_r}^2 / g_i^2} + \frac{1}{\sigma_q^2(x_i)} \right)^{-1}$$

and the local estimate for the conditional entropy is

$$H(x|r(x_i)) \approx \tfrac{1}{2} \ln \left(2\pi e \sigma_{x|r}^2(x_i) \right)$$

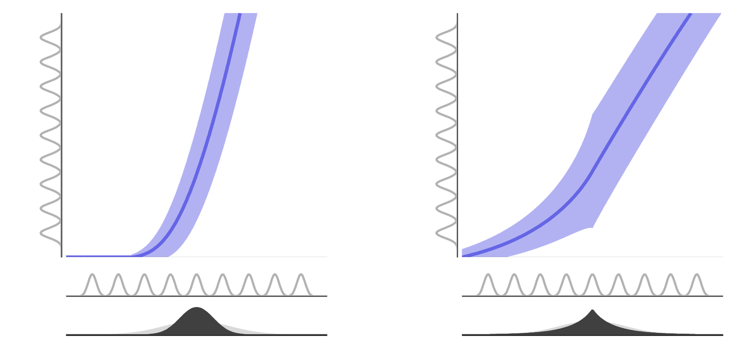
effect of noise and input distribution on the optimal nonlinearity



nonlinearity for a Gaussian input and set noise levels

input and output noise levels affect shape of nonlinearity as well as variability of response

metabolic cost (per spike) adjusted to keep constant mean response



the effect of input pdf on nonlinearity

analytic results for the optimal nonlinearity

we derive the optimal nonlinearity in closed form, given
1. known input distribution *p(x)*
2. additive Gaussian input noise of variance $\sigma_{n_x}^2$
3. additive Gaussian output noise of variance $\sigma_{n_r}^2$
4. metabolic cost linear with firing rate

$$E_x \left[-H(x|r) - \lambda r \right] \propto \int p(x) \left[\log \left(\frac{1}{\sigma_{n_x}^2 + \sigma_{n_r}^2 / g(x)^2} + \frac{1}{\sigma_q^2} \right) - \lambda f(x) \right] dx$$

reparameterize quantities of interest and solve the *discretized* problem:

$$\{h_n\}^* = \arg\max \sum_n \left[z_n \log \left(\frac{1}{1 + h_n^{-2}} + \alpha^2 \right) - \beta h_n - \mu_n h_n \right] \Delta x_n$$

the solution is parameterized by

$h_n = g_n \sigma_{n_x} / \sigma_{n_r}$ (slope of nonlinearity scaled by ratio of input/output noise)

and depends on

$\alpha = \sigma_{n_x} / \sigma_x$ (input SNR)

$\beta = \lambda \sigma_{n_r} / \sigma_{n_x}$ (ratio of input/ouput noise)

$z_n = p_n / (1 - P_n)$ (ratio of input pdf to 1-cdf, also $= -\frac{d}{dx} \log(1 - P(x))$)

μ_n (Lagrange multipliers for $h_n \geq 0$ constraint)

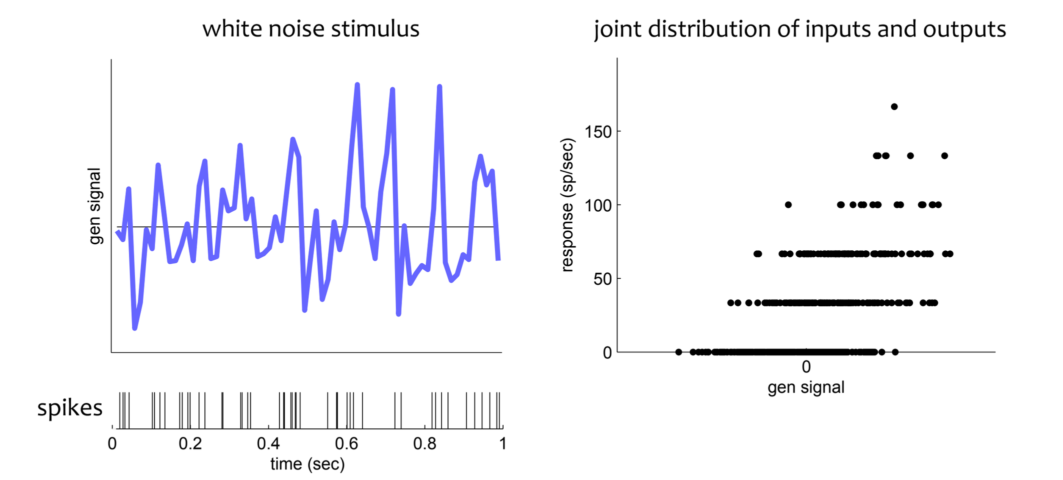
the optimal nonlinearity is given by the solution to the fourth-order polynomial

$$\frac{dL}{dh_n} = -\frac{2 h_n z_n}{(1 + h_n^2)(\alpha^2 + \alpha^2 h_n^2 + h_n^2)} - \beta - \mu_n$$

(not pretty, but can be computed in closed form)

observations
- the solution (up to scaling) lives in a 2D space defined by α, β
- spike cost and firing rate precision trade off (through β)
- data probability density affects solution through z_n (the cumulative hazard function)
- when $\mu_n \neq 0$, hard thresholding rectification
- high input SNR implies $\alpha^2 \to \infty$, solution is off the constraint boundary, $\mu_n = 0$, no hard rectification
  this leads to vanishing noise results (Laughlin, 1981; Nadal and Parga, 1994)
- we can incorporate level-dependent input noise using bin-specific values α_n, β_n
- level dependent neural noise requires iterative computation of slope variables h_n

analysis of retinal ganglion cell data



data: in vitro RGC responses (Chichilnisky and Kalmar, 2002)

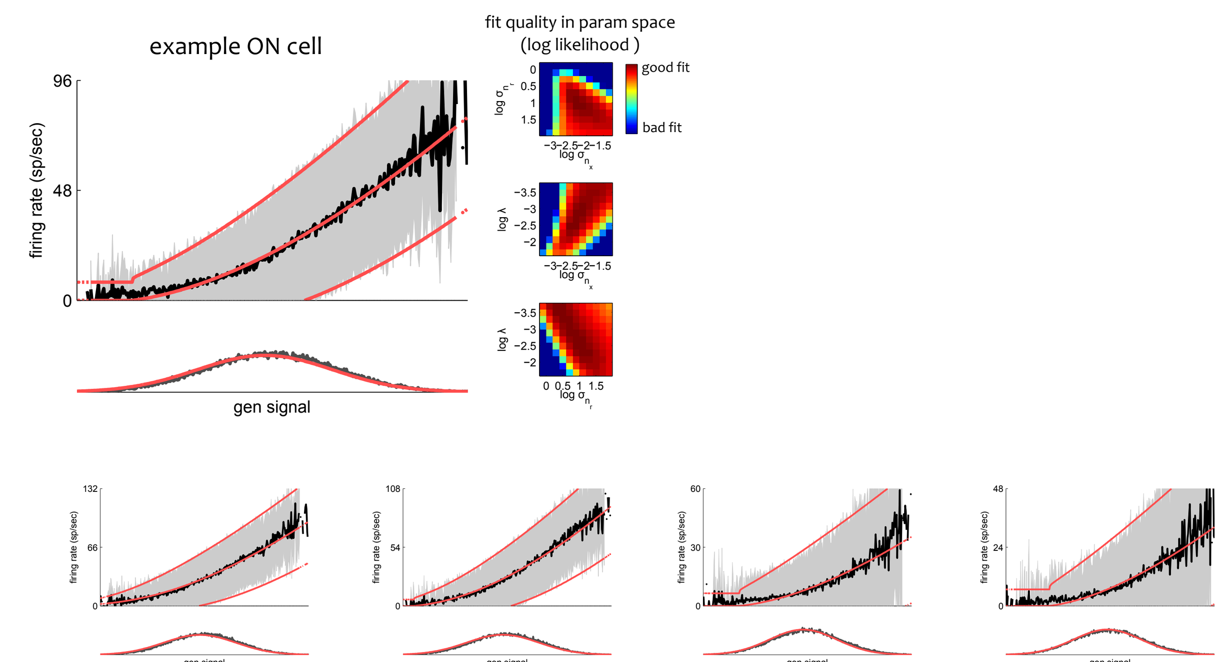
input: binary white noise stimulus convolved with spatiotemporal STA (generator signal)

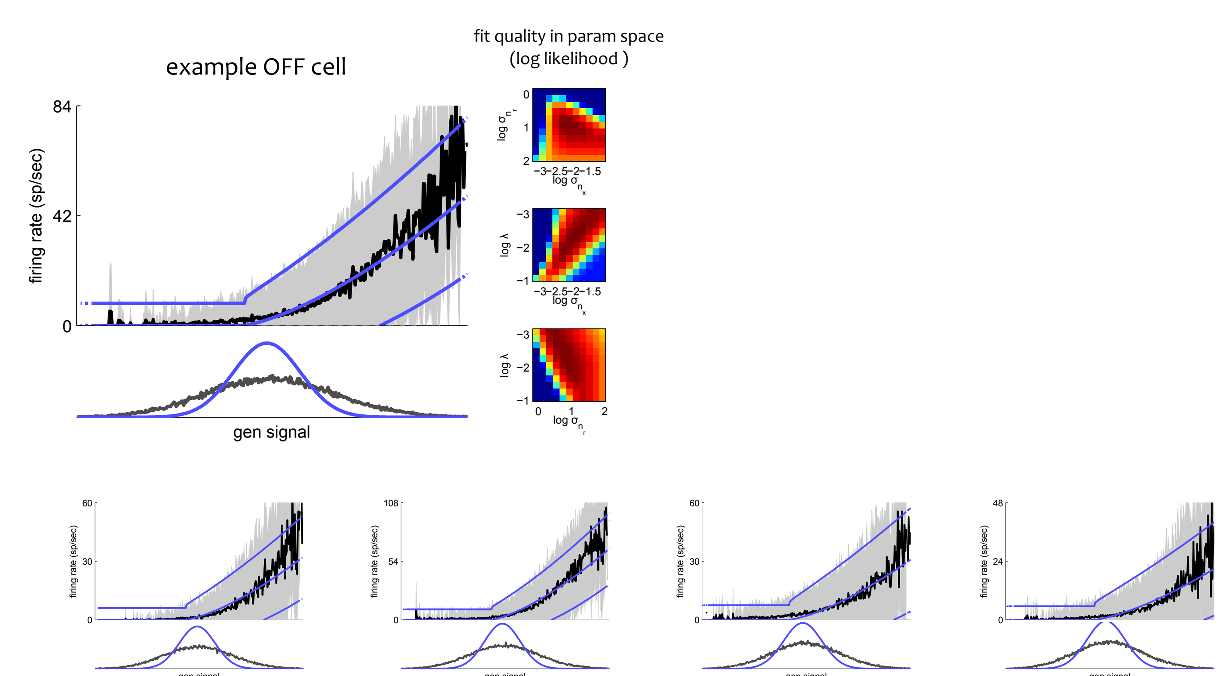
output: number of spikes in each time frame (instantaneous firing rate)

response mean and s.d.: computed for 500 (binned) input values

fitting methods

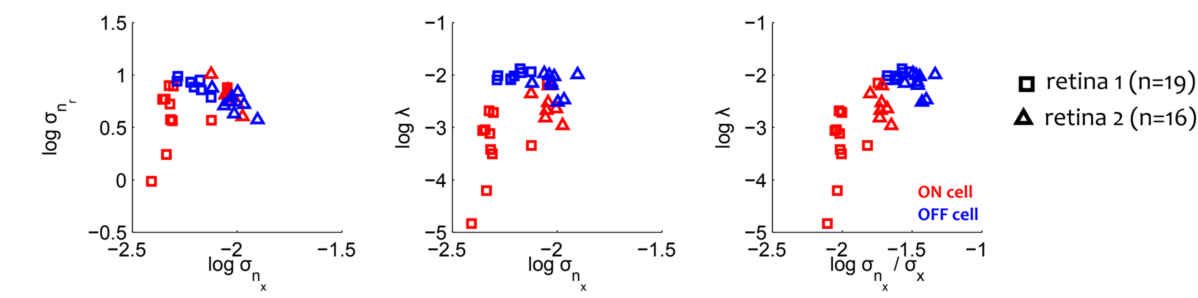
- search over $\sigma_{n_x}^2, \sigma_{n_r}^2, \lambda$ for nonlinearity and predicted variability that best match input-output pairs (maximize data likelihood under Gaussian noise model)
- also estimate optimal input distribution (generalized Gaussian family: σ_x, μ_x, d_x) jointly over all ON cells and jointly over all OFF cells





population analysis

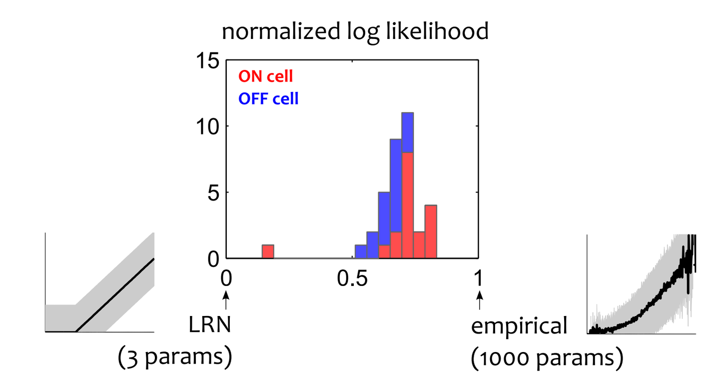
what is the distribution of inferred parameters?



quantitative model comparison

how well does the model fit the data?

compare the likelihood under the infomax model (~3 params) to two descriptive models
1. Linear + Rectify + fixed output Noise
2. empirical (mean and std dev in each bin)



summary

contributions

we extend previous Infomax results that relied on *vanishing noise*, and derive the nonlinearity that maximizes MI in the presence of non-negligible input and output noise and metabolic cost (in bits) that is proportional to firing rate

in contrast to the vanishing noise case, hard rectification (i.e. zero mean firing rate for inputs below some threshold) is optimal for some parameter settings

the model provides good fits to retinal data; makes predictions about noise levels, metabolic cost parameters, and adapted input distributions; optimal input pdf (fit exponent) is nearly Gaussian

future work

- *theoretical*: approximation error when noise is significant; level-dependent noise)

- *applications*: Bayesian noise estimates; adaptive behavior during stimulus changes

- *model extensions*: linear filter (nD input); joint encoding by multiple neurons (generalize ICA)

related theoretical work

long history of nonlinear models that maximize information or reconstruction quality of stimulus

many incorporate elements of current work

	$f(x)$	$p(r)$	σ_{n_x}	σ_{n_r}	metabolic constr	obj
Laughlin, 1981	✓	✓	×	$\sigma_{n_r} \to 0$	r_{max}	MI
Schuster, 1992	✓	×	if f linear	✓	r_{max}	MI
Nadal and Parga, 1994	✓	×	$\sigma_{n_x} < \epsilon$	$\sigma_{n_r} \to 0$	r_{max}, other	MI
Treves et al, 1999	×	✓	×	×	fix $E[r]$	MI
von der Twer and MacLeod, 2001	✓	×	✓	✓	r_{max}	MSE
Balasubramanian et al, 2001; de Polavieja 2002	×	✓	✓	✓(discrete)	$Cost[r_k]$	MI
McDonnell and Stocks, 2008	✓	×	✓	✓	r_{max}	MI