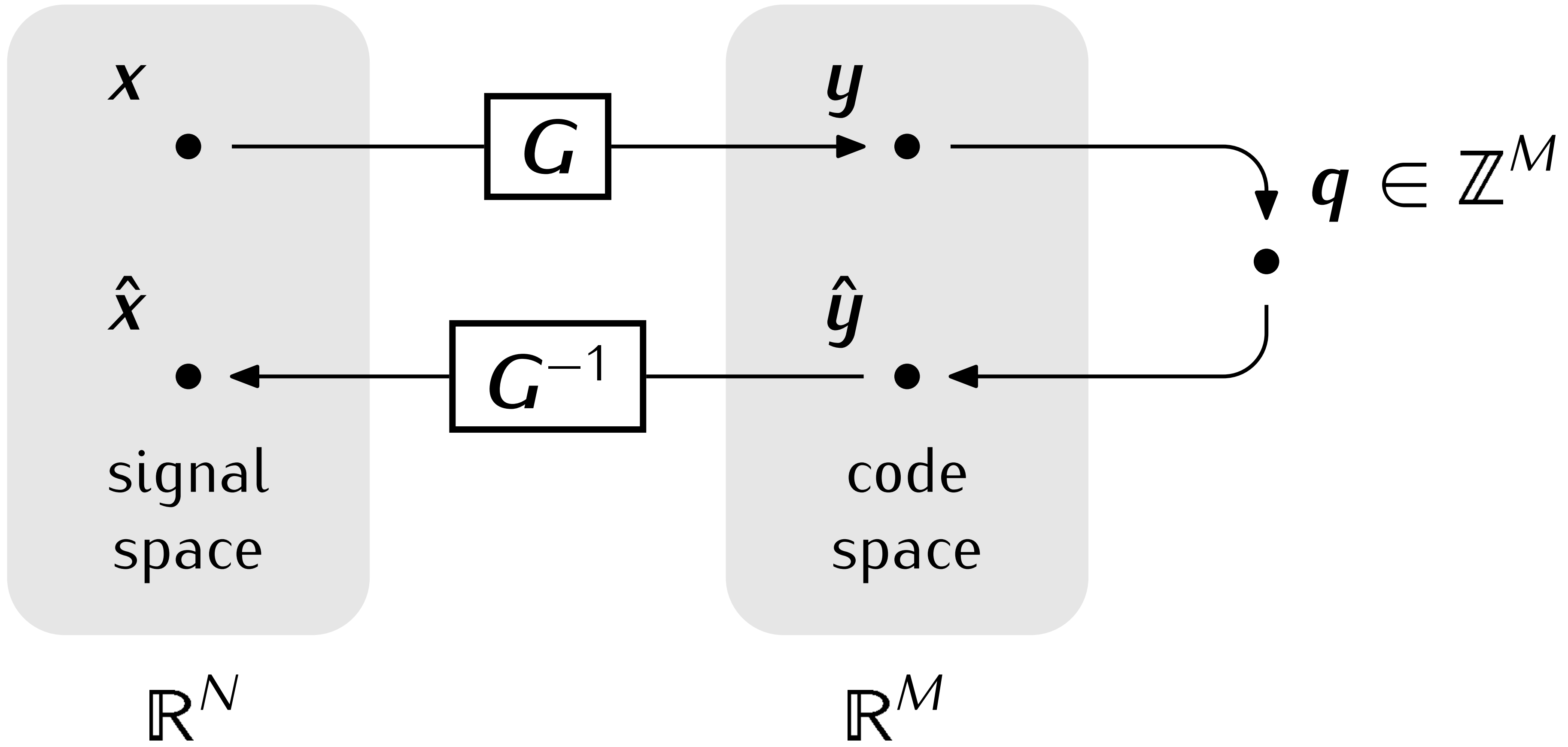
End-to-end optimized image compression

Johannes Ballé
Center for Neural Science, NYU
Howard Hughes Medical Institute
(now with Google Inc.)

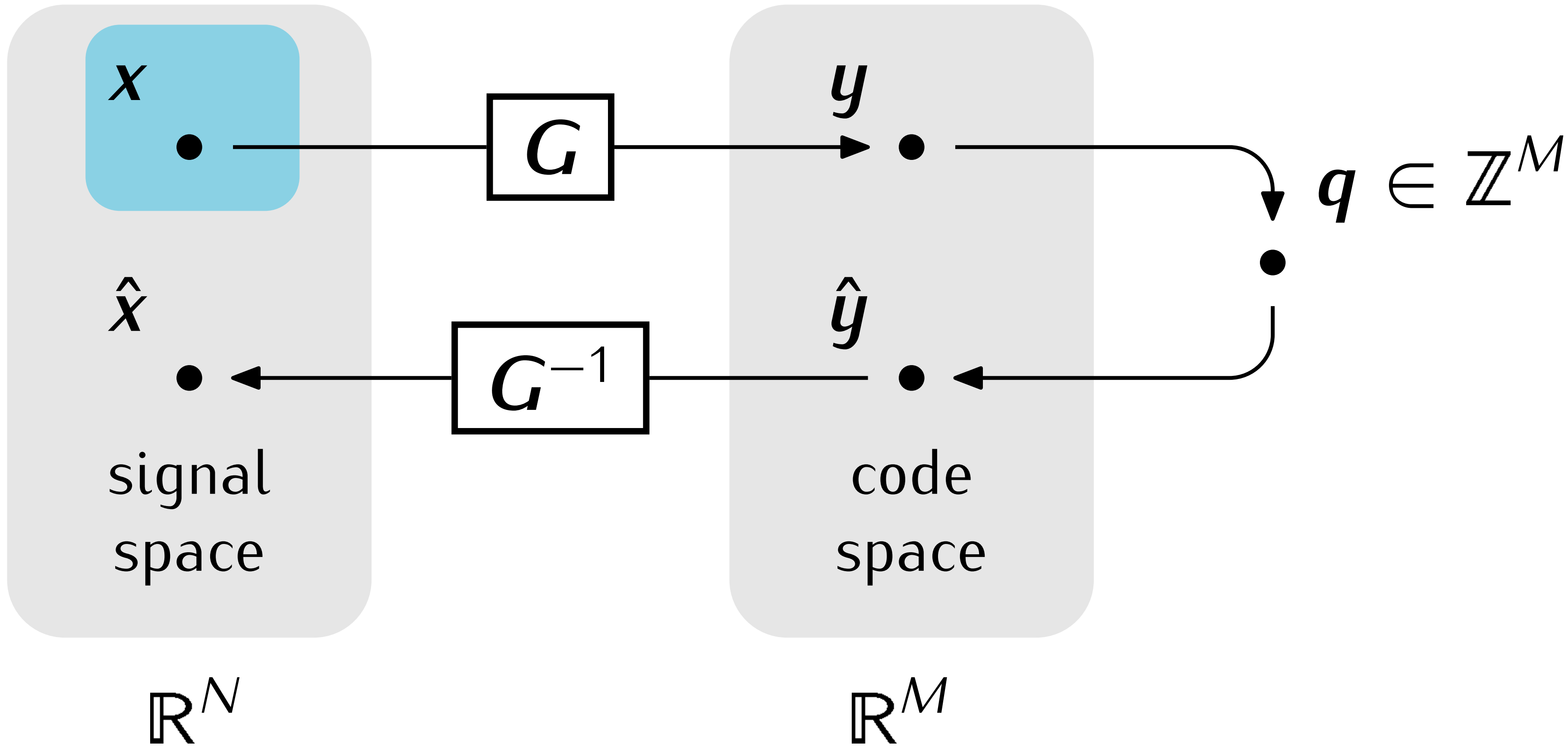
joint work with:
Valero Laparra, Universitat de València
Eero P. Simoncelli, CNS/Courant Institute/HHMI

International Conference on Learning Representations, Toulon, 2017

Linear transform coding

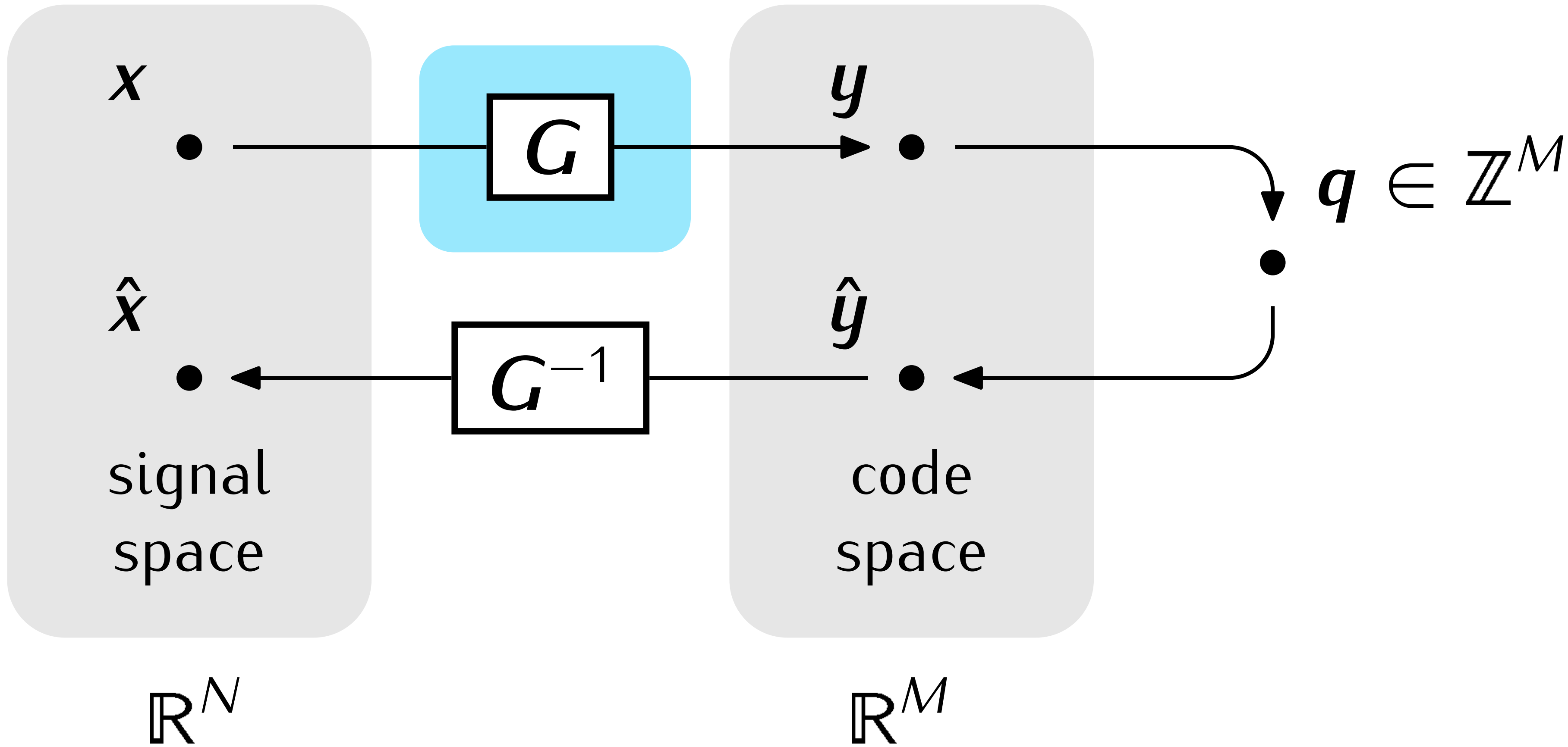


Linear transform coding

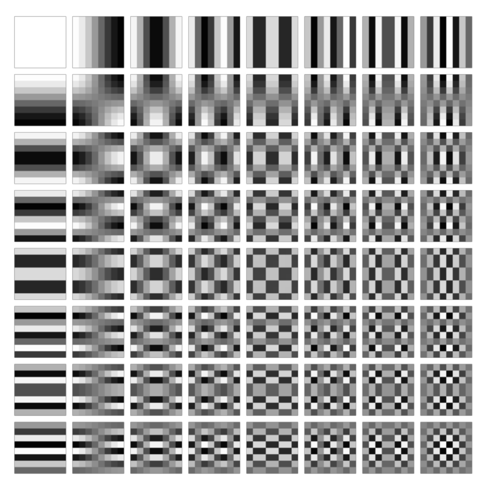


x

Linear transform coding

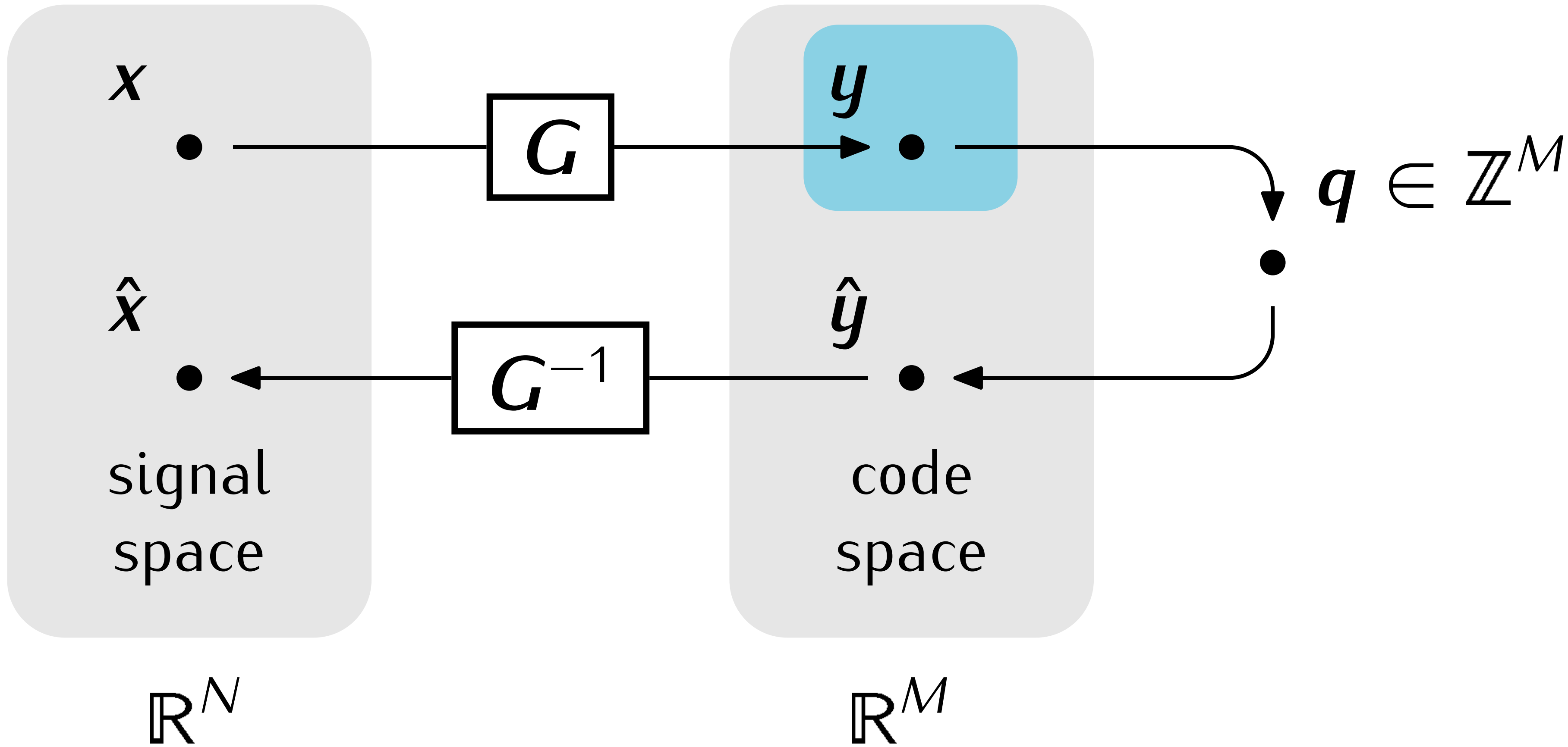


x

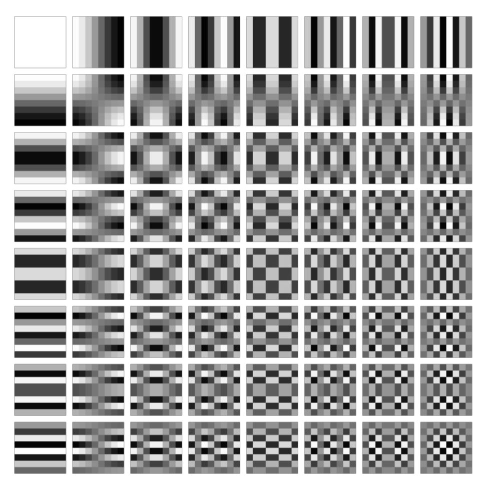


G

Linear transform coding

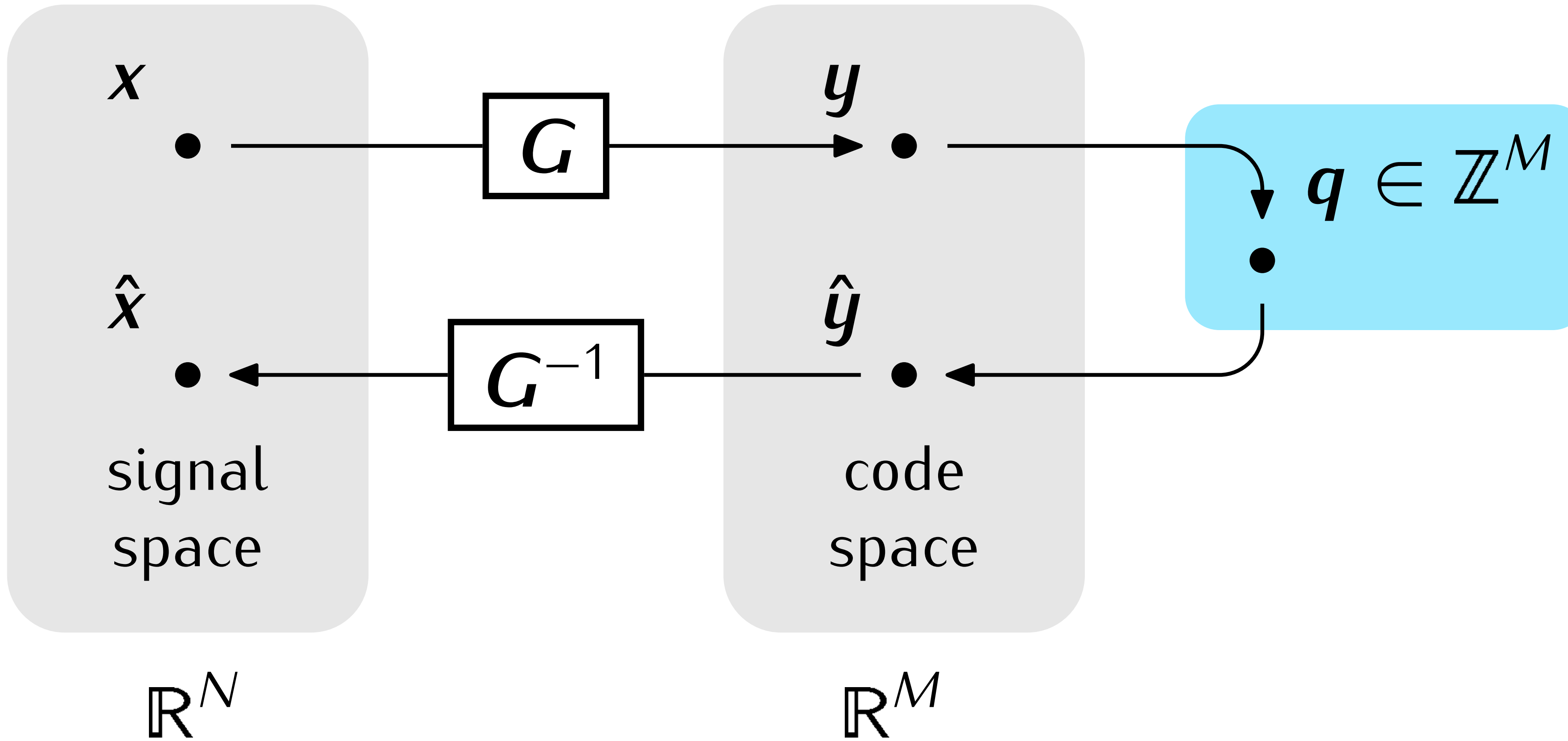


x

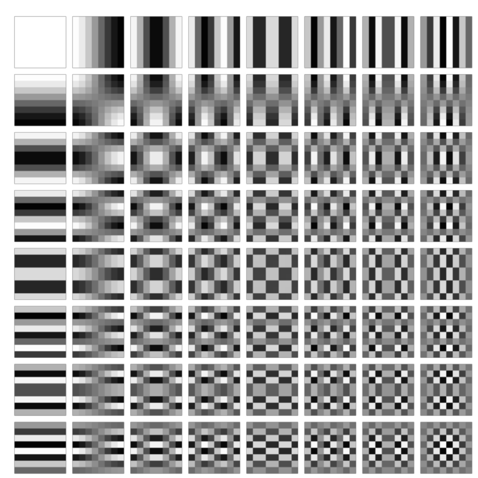


G

Linear transform coding

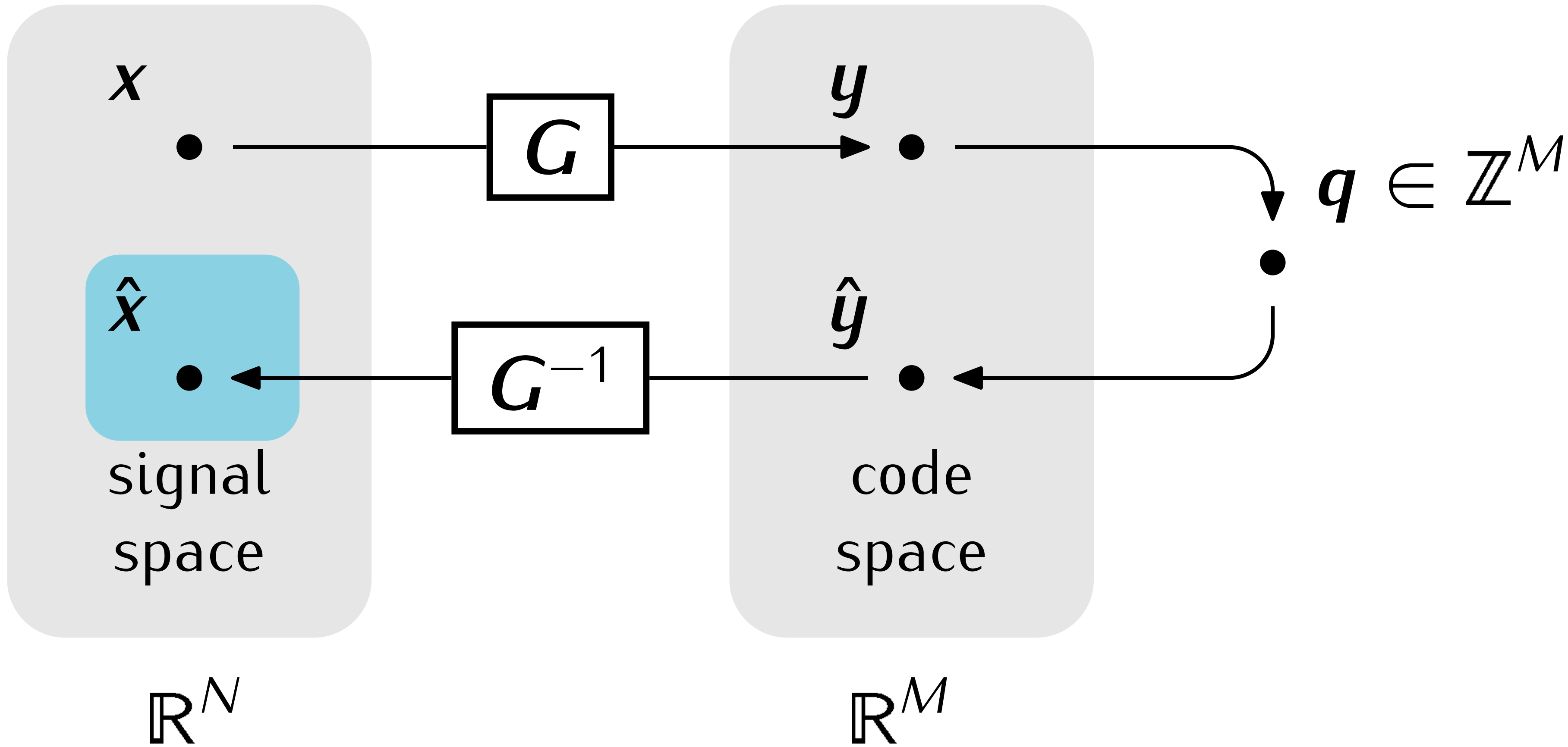


$\boldsymbol{x}$

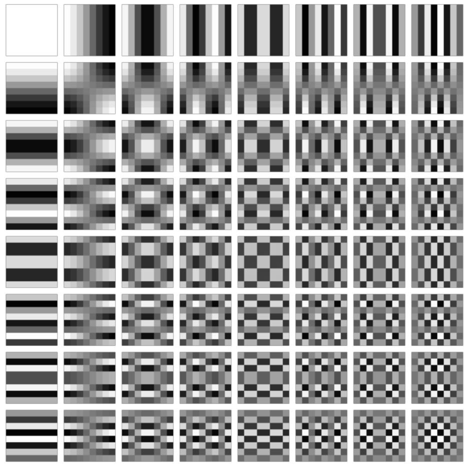


G

Linear transform coding



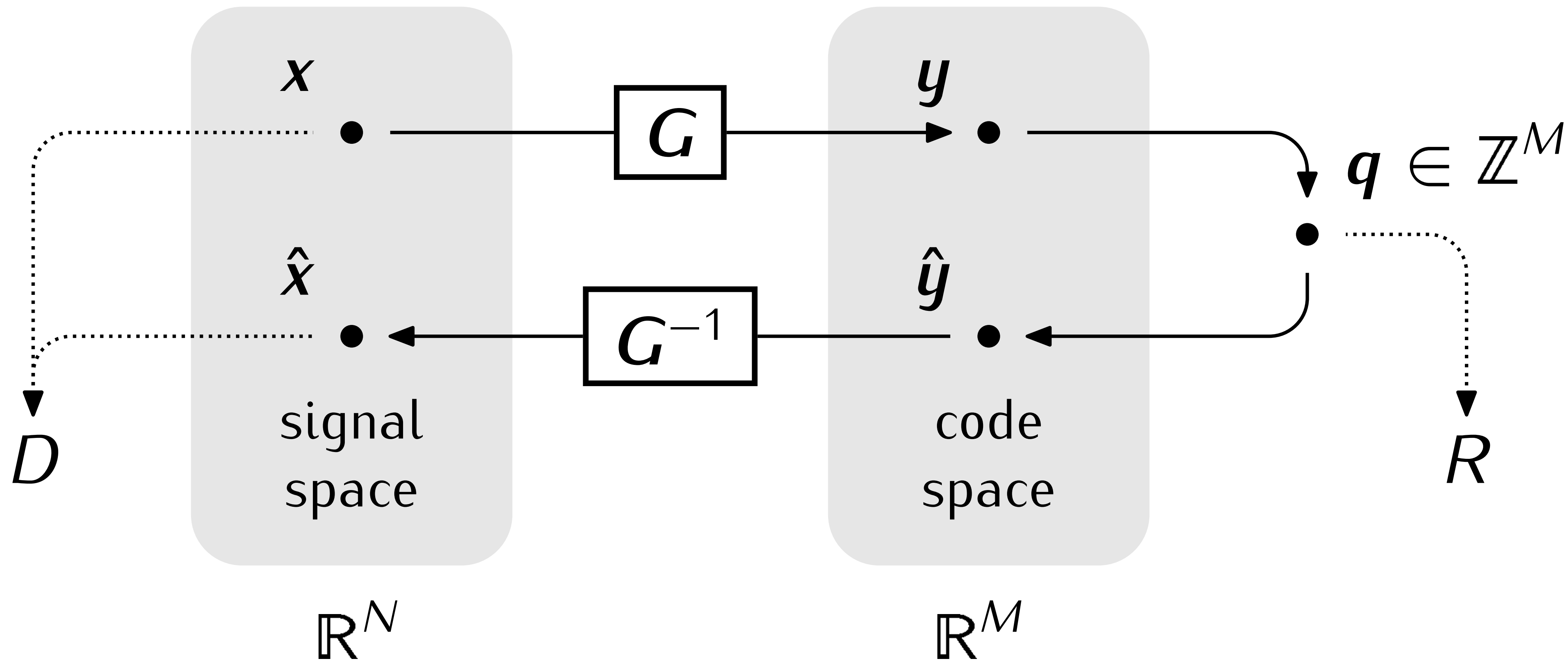
$\boldsymbol{x}$



G

$\hat{\boldsymbol{x}}$

Linear transform coding



D: distortion, e.g. mean squared error

R: rate, ideally close to Shannon entropy of $\boldsymbol{q}$

rate: 0.17 bits/pixel

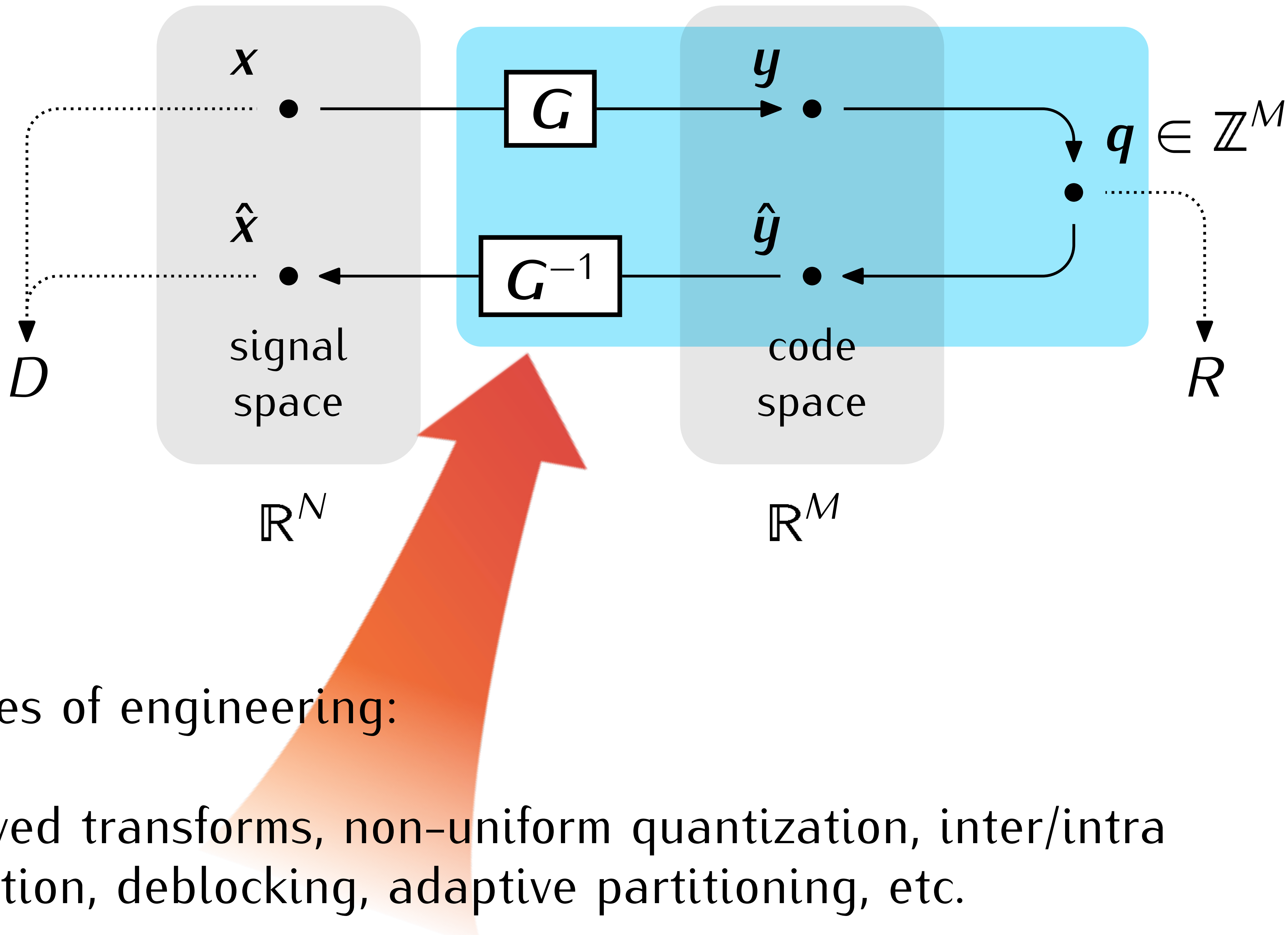
rate: 0.12 bits/pixel

coarser quantization: lower rate, higher distortion

rate: 0.32 bits/pixel

finer quantization: higher rate, lower distortion

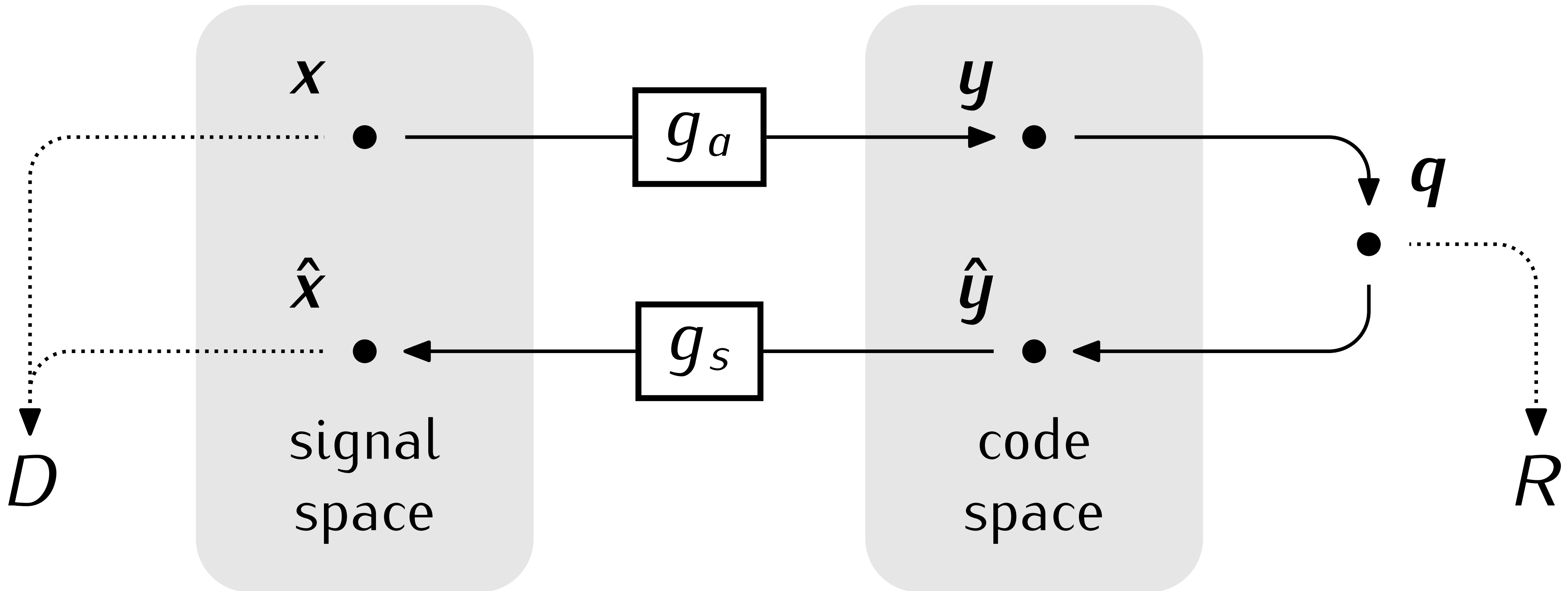
Linear transform coding



decades of engineering:

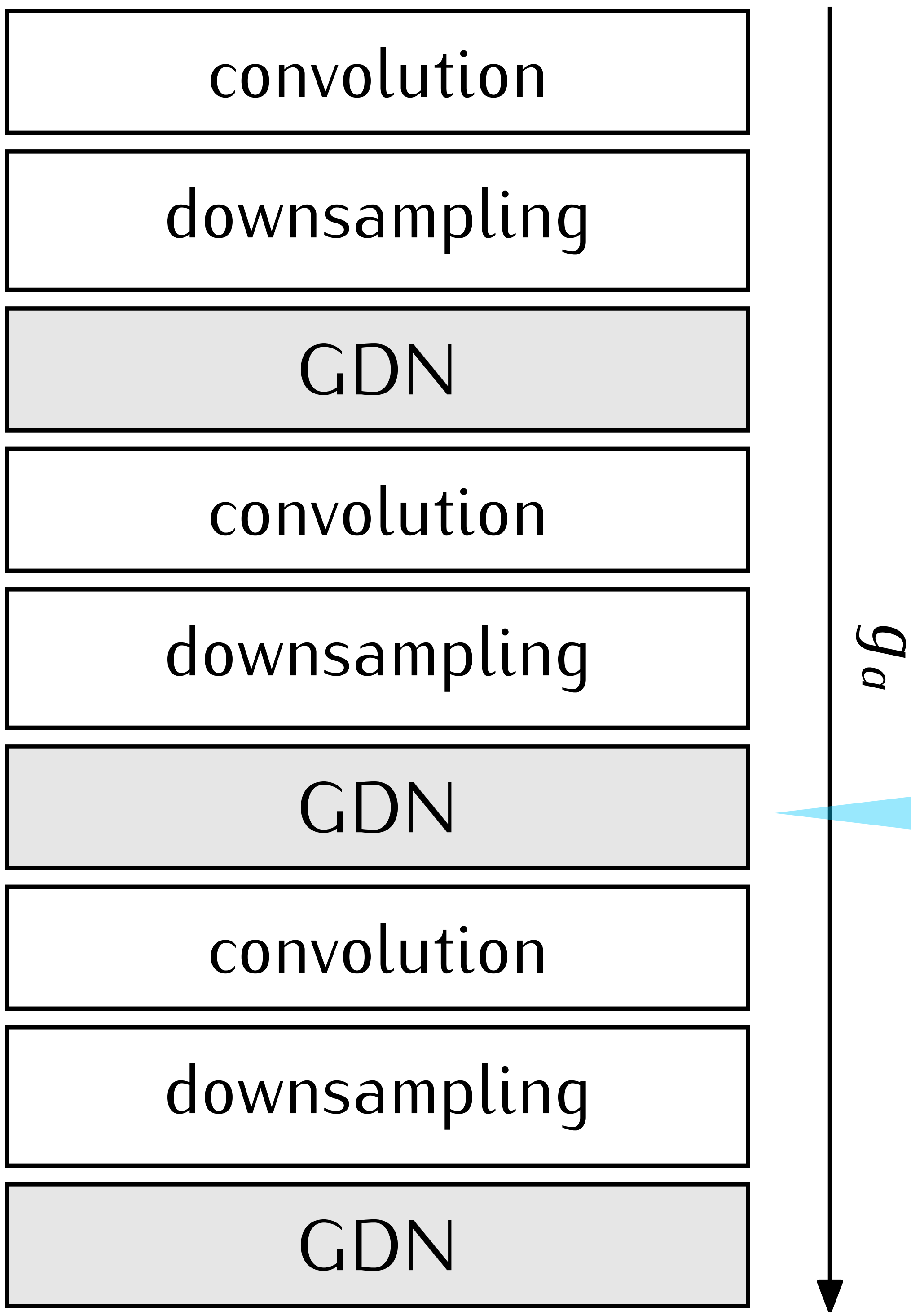
improved transforms, non-uniform quantization, inter/intra prediction, deblocking, adaptive partitioning, etc.

Nonlinear transform coding



g_a, g_s: multivariate, parametric nonlinear functions
(if it helps, think of them as neural networks)

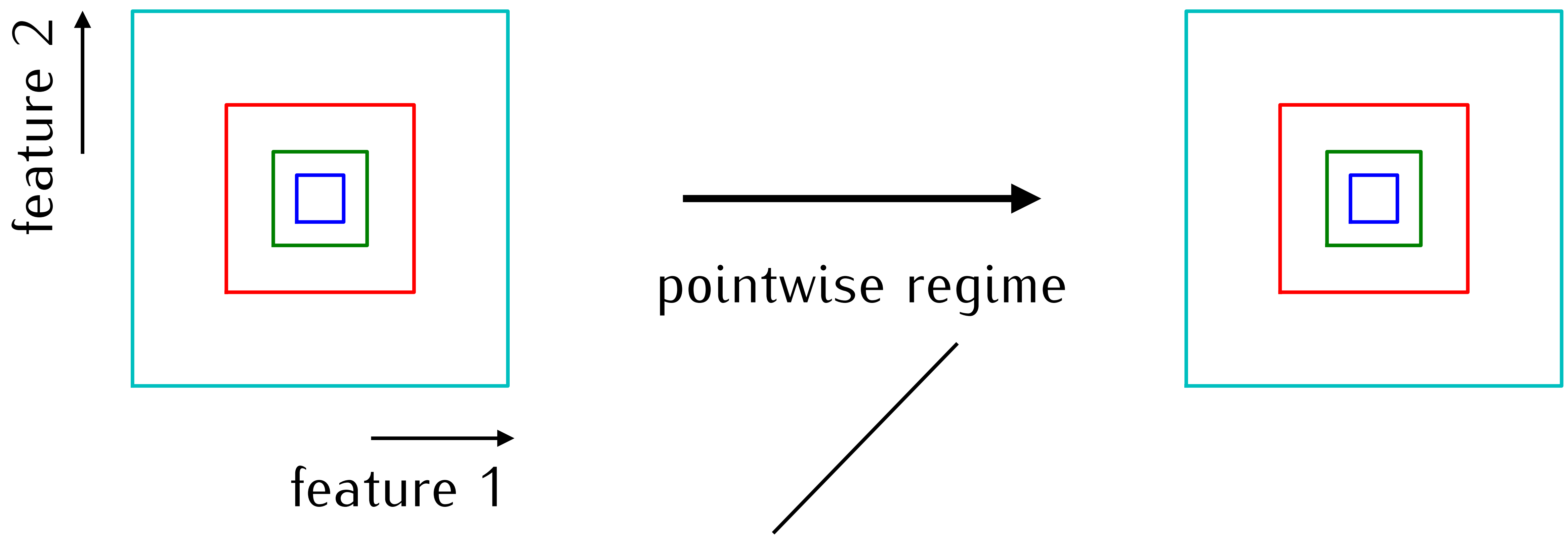
Architecture of transformation



generalized divisive normalization

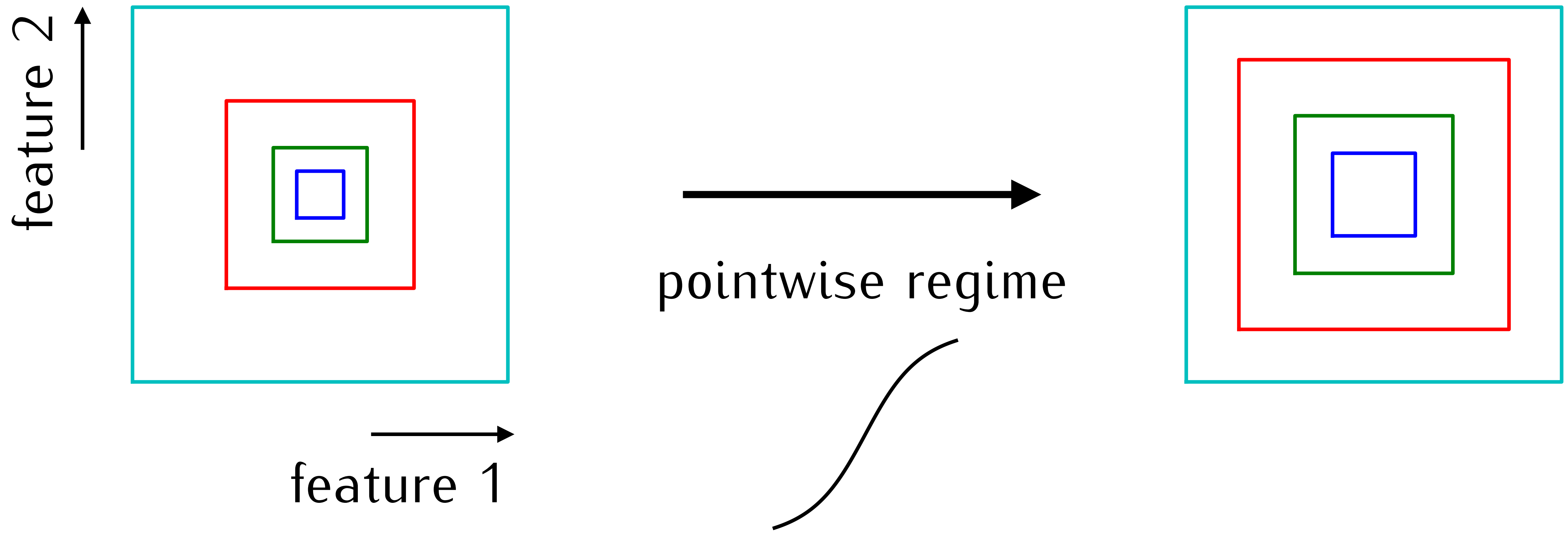
$$u_i = \frac{w_i}{\sqrt{\beta_i + \sum_j \gamma_{ij} |w_j|^2}}$$

Generalized divisive normalization (GDN)



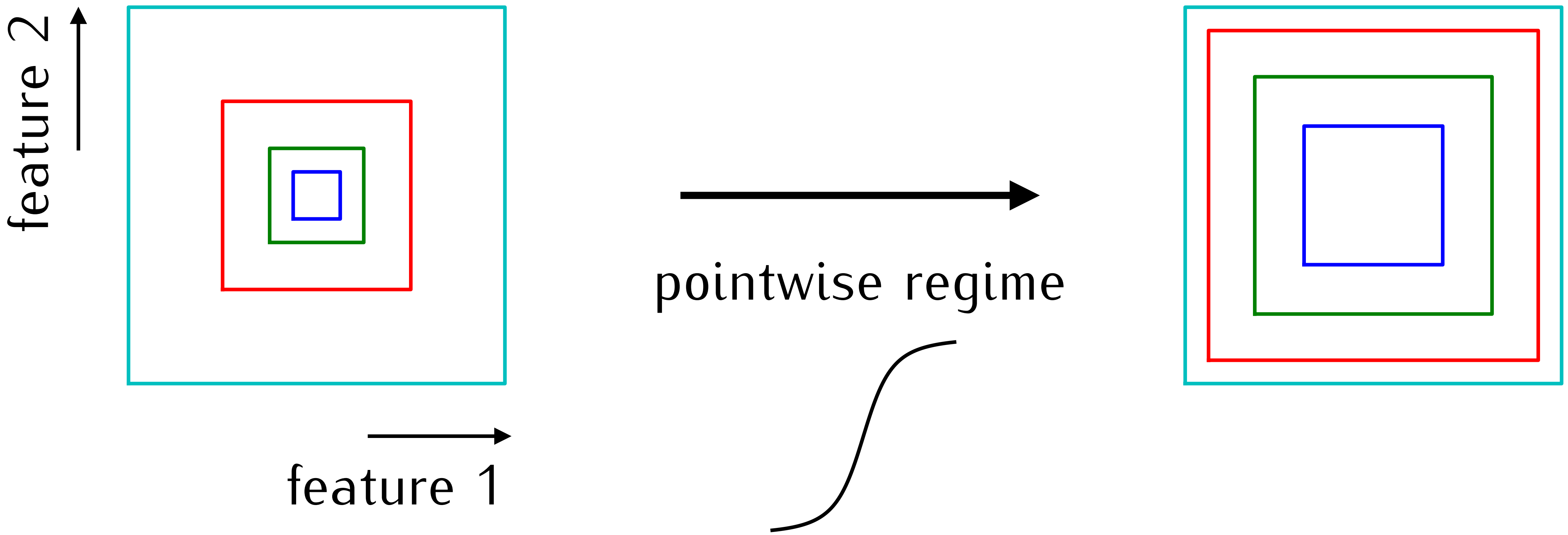
generalization of:
– sigmoid-type nonlinearities
– local response normalization (LRN)
see our ICLR 2016 paper for details

Generalized divisive normalization (GDN)



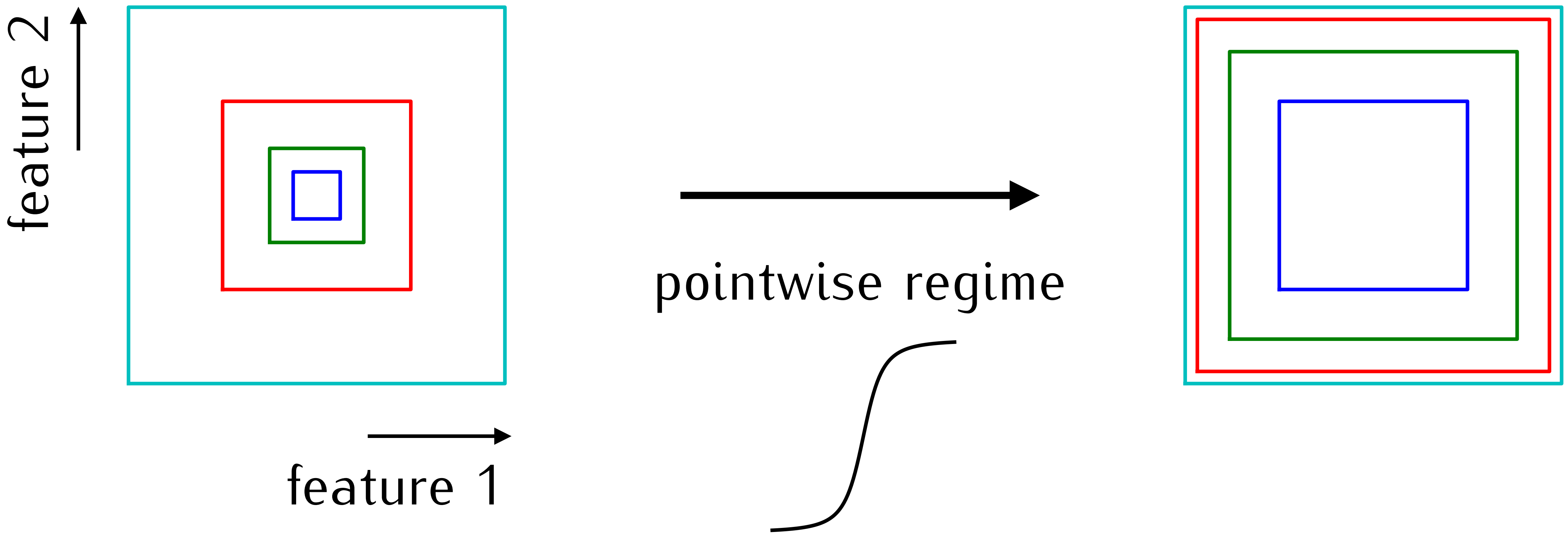
generalization of:
– sigmoid-type nonlinearities
– local response normalization (LRN)
see our ICLR 2016 paper for details

Generalized divisive normalization (GDN)



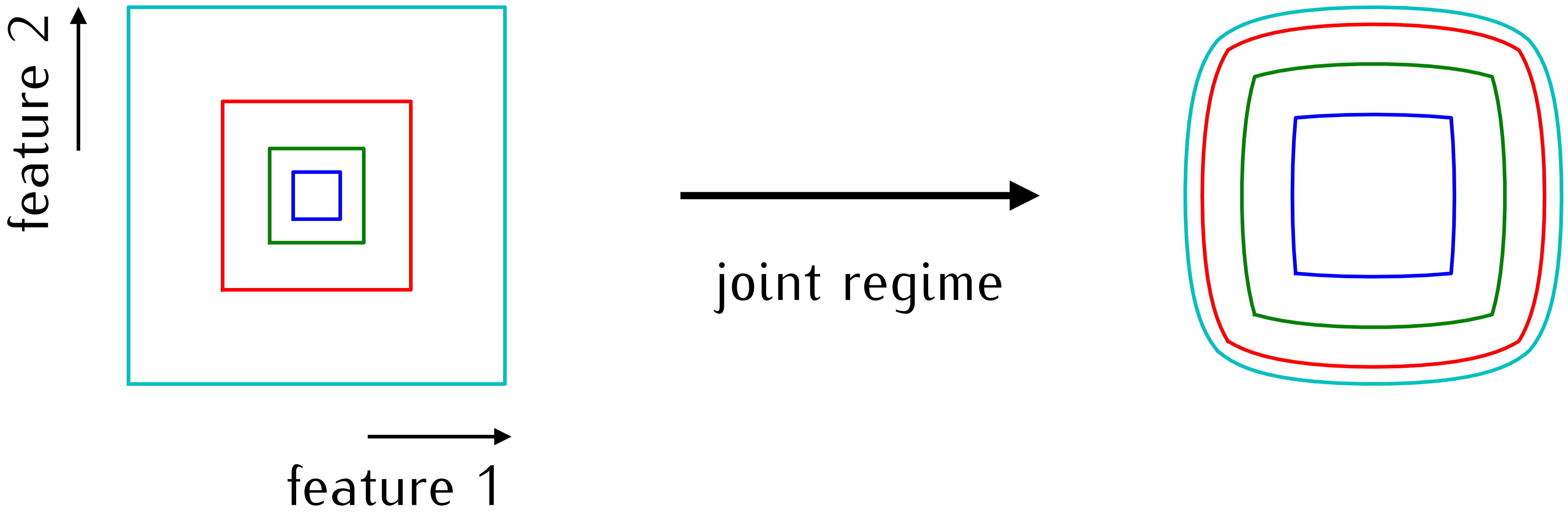
generalization of:
– sigmoid-type nonlinearities
– local response normalization (LRN)
see our ICLR 2016 paper for details

Generalized divisive normalization (GDN)



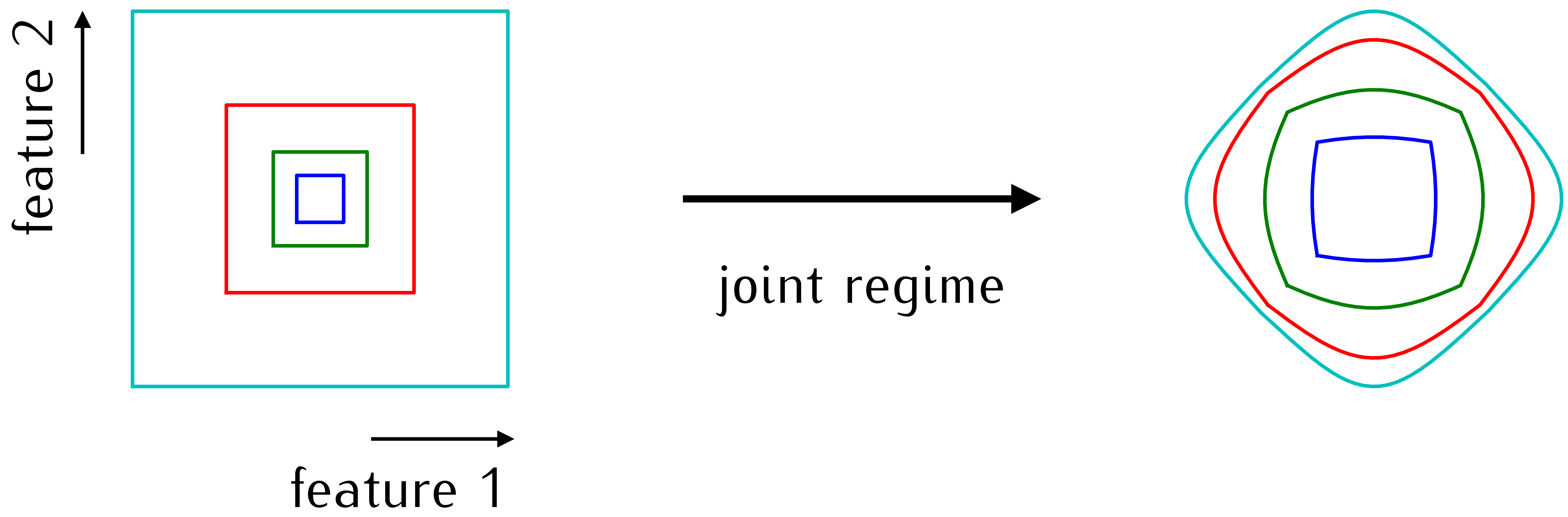
generalization of:
– sigmoid-type nonlinearities
– local response normalization (LRN)
see our ICLR 2016 paper for details

Generalized divisive normalization (GDN)



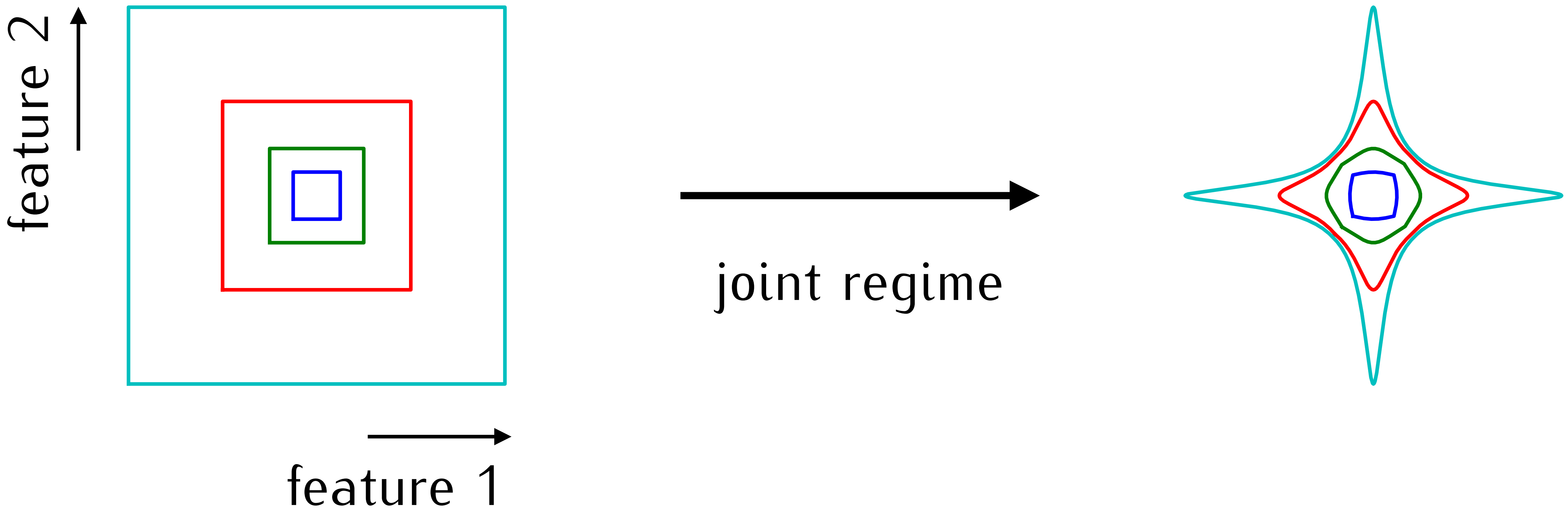
generalization of:
– sigmoid-type nonlinearities
– local response normalization (LRN)
see our ICLR 2016 paper for details

Generalized divisive normalization (GDN)



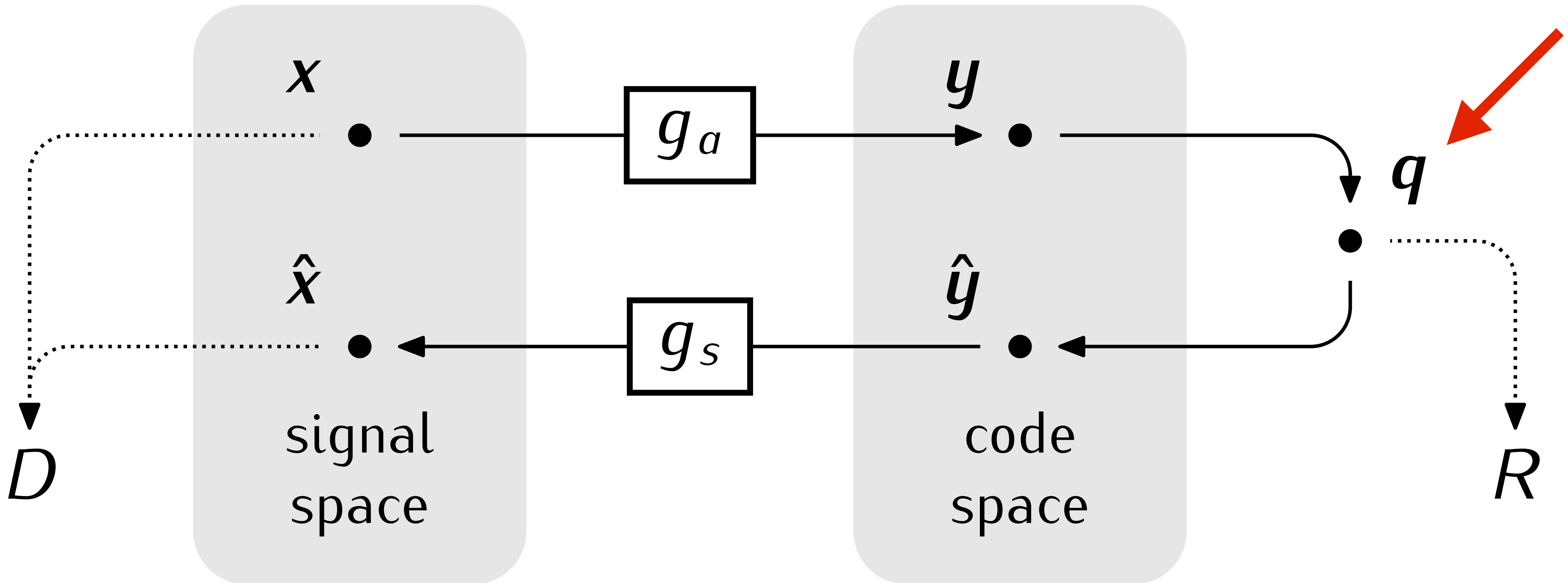
generalization of:
– sigmoid-type nonlinearities
– local response normalization (LRN)
see our ICLR 2016 paper for details

Generalized divisive normalization (GDN)



generalization of:
– sigmoid-type nonlinearities
– local response normalization (LRN)
see our ICLR 2016 paper for details

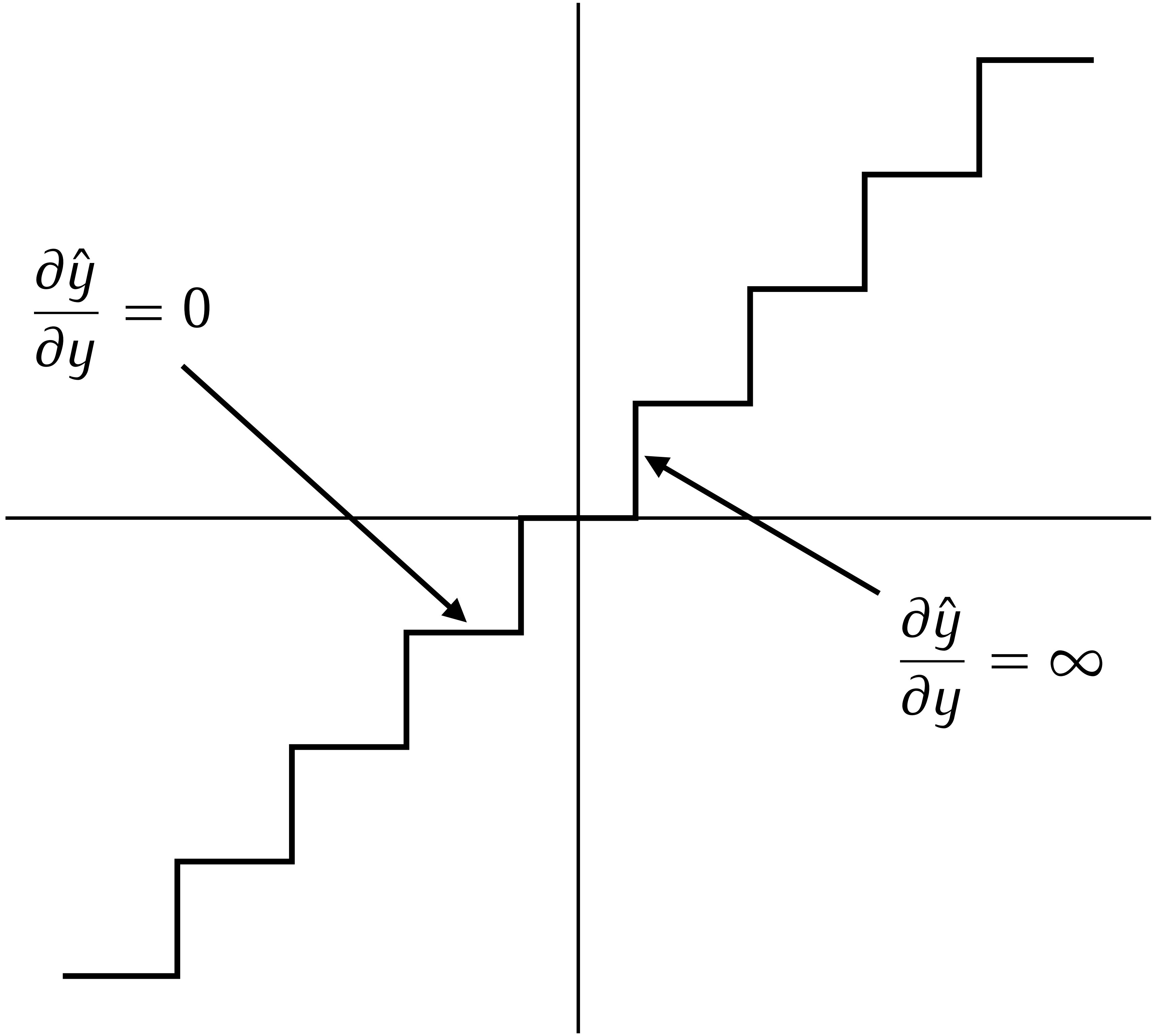
Nonlinear transform coding

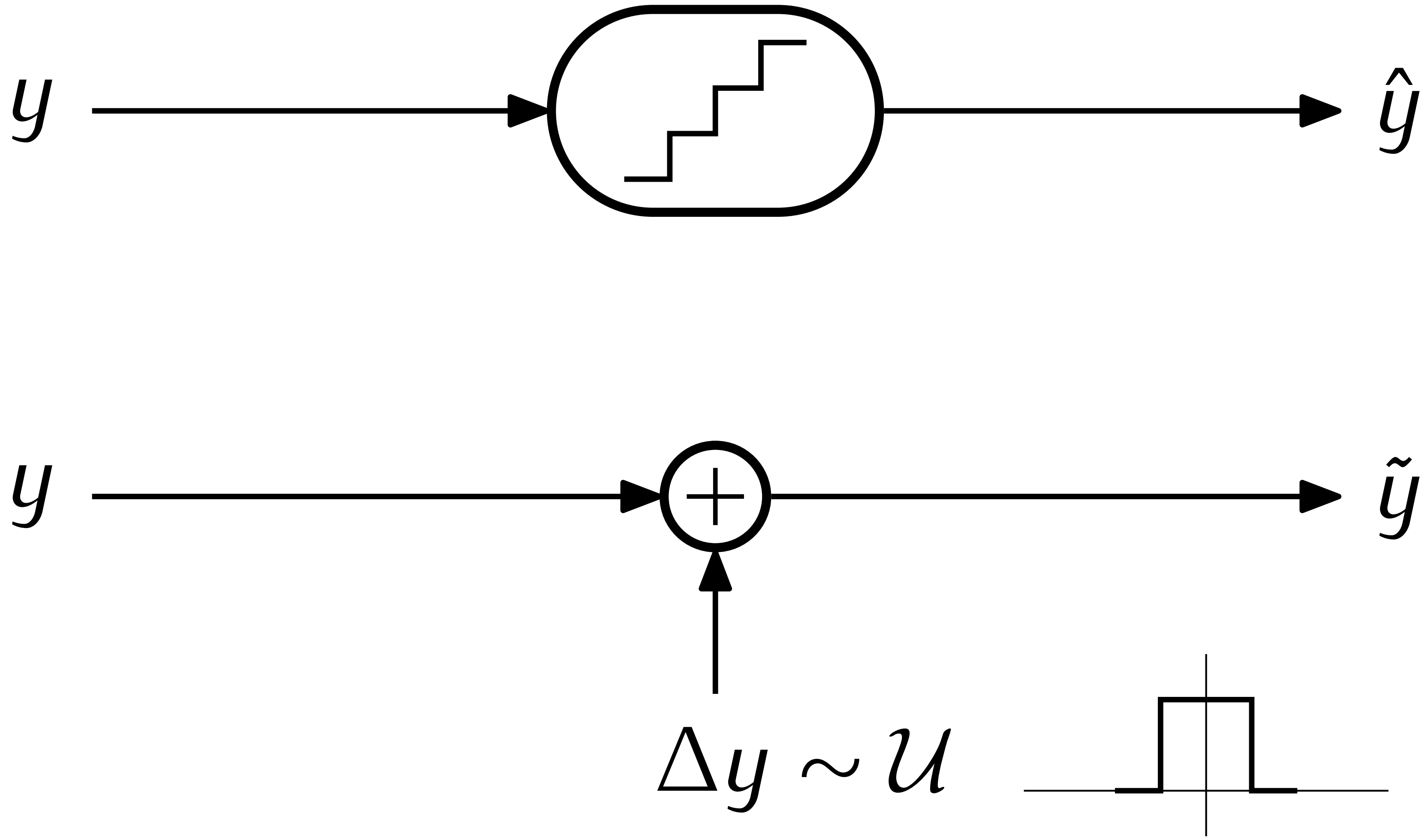


optimize g_a, g_s for rate and distortion numerically

$$L[g_a, g_s, P_{\boldsymbol{q}}] = \underbrace{-\,\mathbb{E}[\log_2 P_{\boldsymbol{q}}]}_{R} + \lambda\,\underbrace{\mathbb{E}[d(\boldsymbol{x}, \hat{\boldsymbol{x}})]}_{D}$$

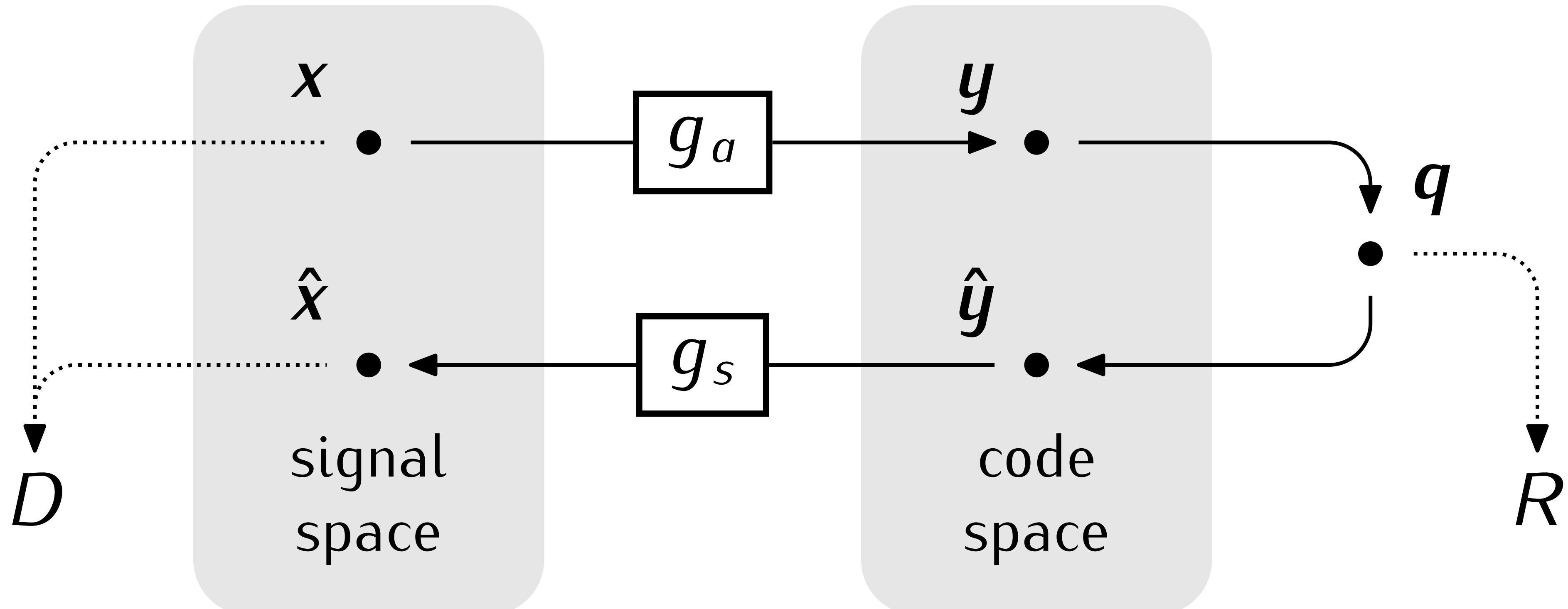
gradient is zero *almost everywhere*



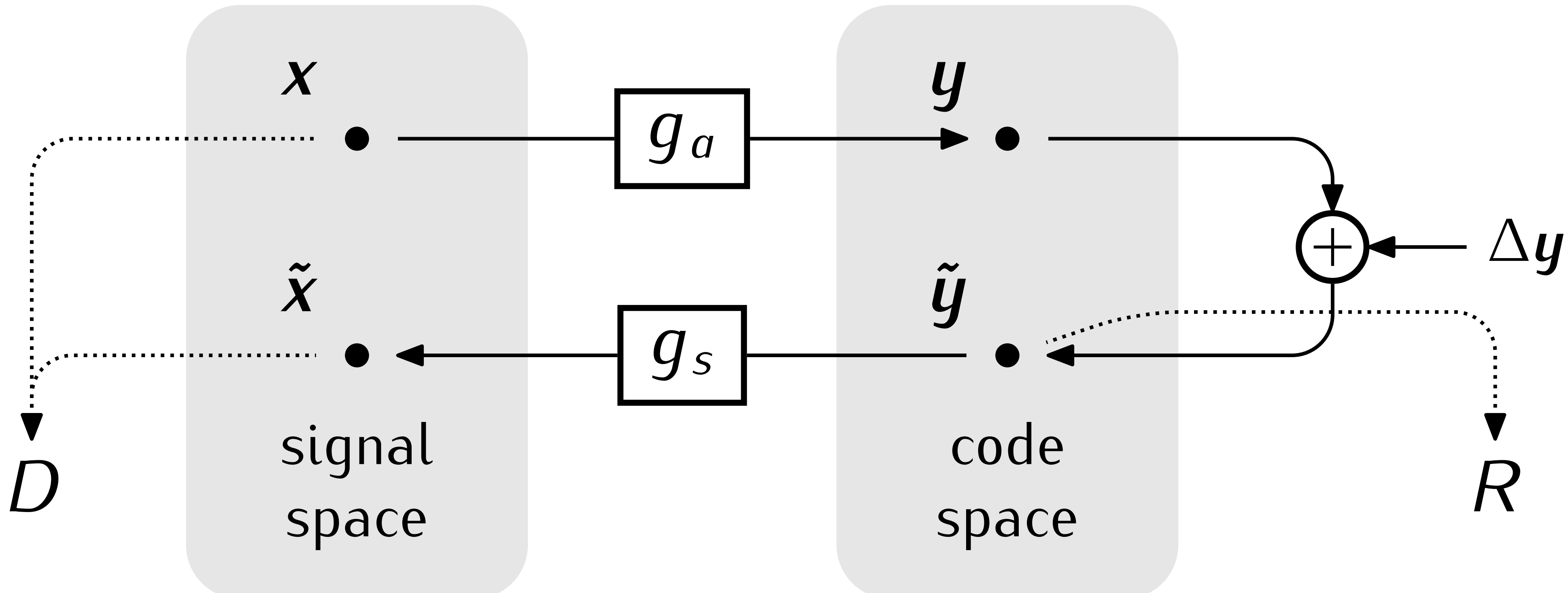


differentiable and continuous
stochastic approximation

other approaches:
Theis et al., 2017
Jang et al., 2017
Maddison et al., 2017



$$L = \mathbb{E}\left[-\sum_i \log_2 P_{q_i}(q_i) + \lambda \left\| \hat{\boldsymbol{x}} - \boldsymbol{x} \right\|_2^2 \right]$$



proxy loss:

$$L = \mathbb{E}\left[-\sum_i \log_2 p_{\tilde{y}_i}(\tilde{y}_i) + \lambda \left\| \tilde{\boldsymbol{x}} - \boldsymbol{x} \right\|_2^2 \right]$$

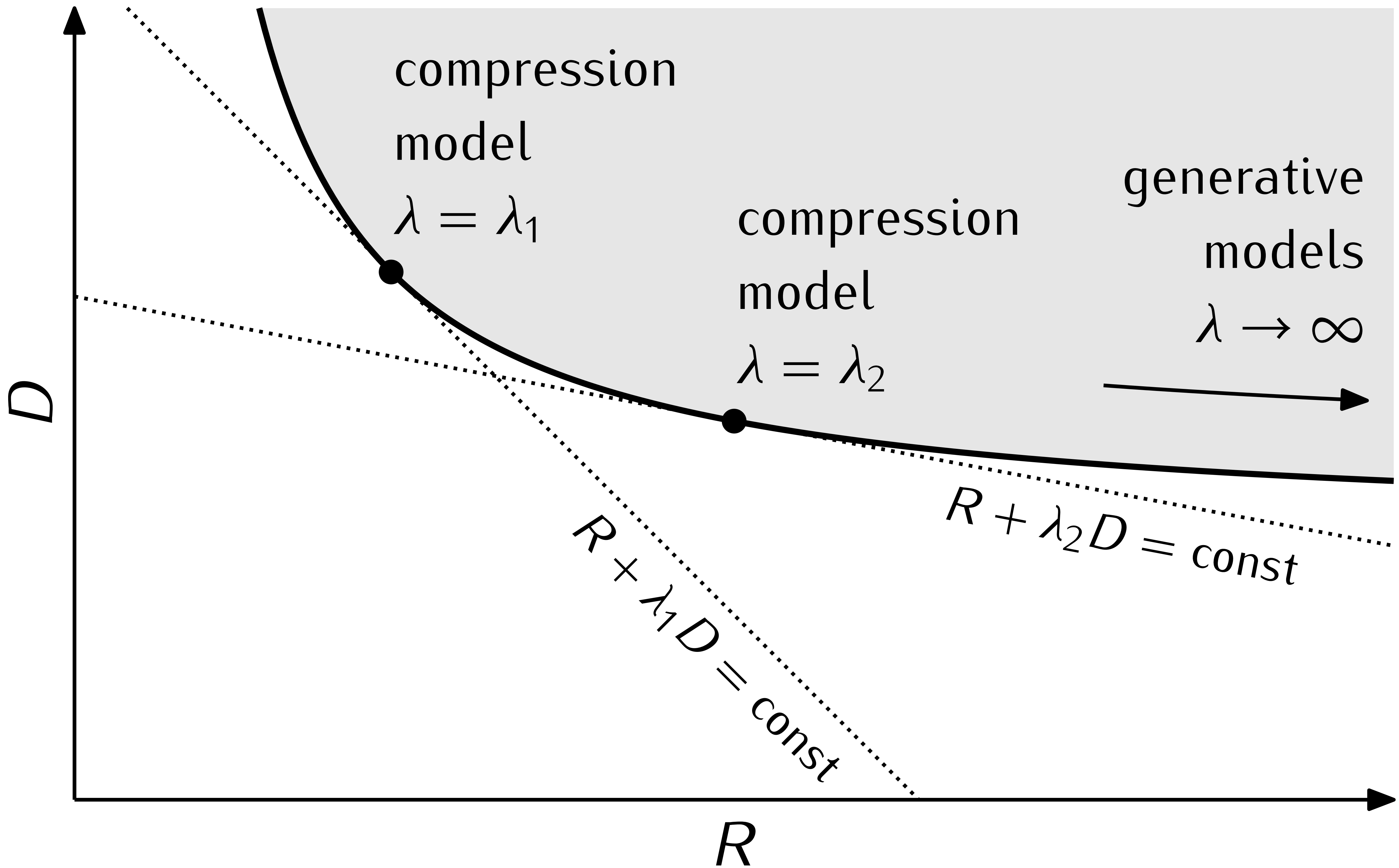
Wait! Isn't this just an autoencoder?

(Yes and no.)

Proxy RD: rate distortion

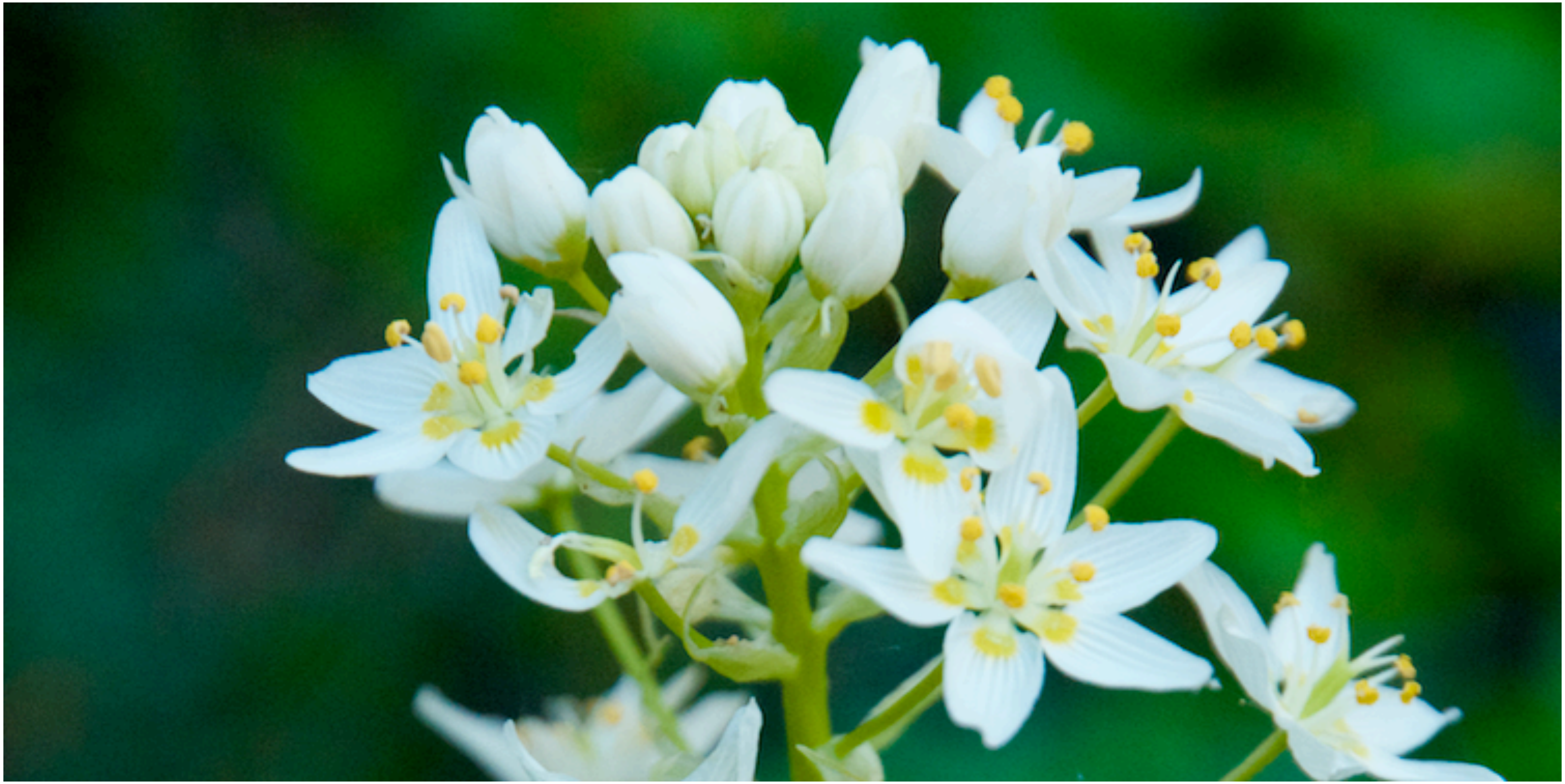
$$L = \mathbb{E}\left[-\sum_i \log_2 p_{\tilde{y}_i}(\tilde{y}_i) + \lambda \left\| \tilde{\boldsymbol{x}} - \boldsymbol{x} \right\|_2^2 \right]$$

Variational AE: log prior log likelihood


compression
model
λ = λ1
compression
model
λ = λ2
generative
models
λ → ∞
R + λ1D = const
R + λ2D = const
D
R

Results

original

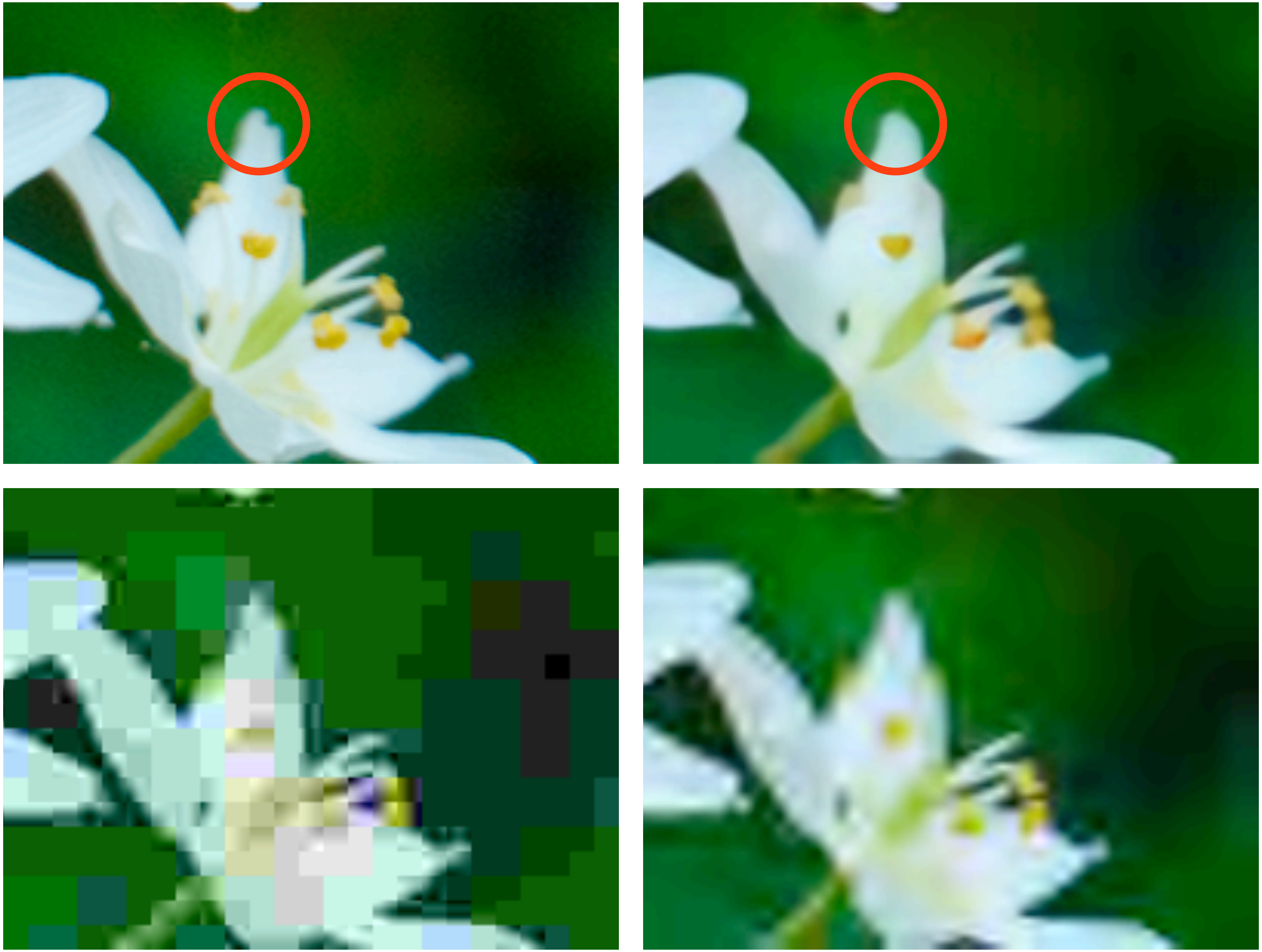


JPEG @ 0.119 bits/px

JPEG 2000 @ 0.107 bits/px

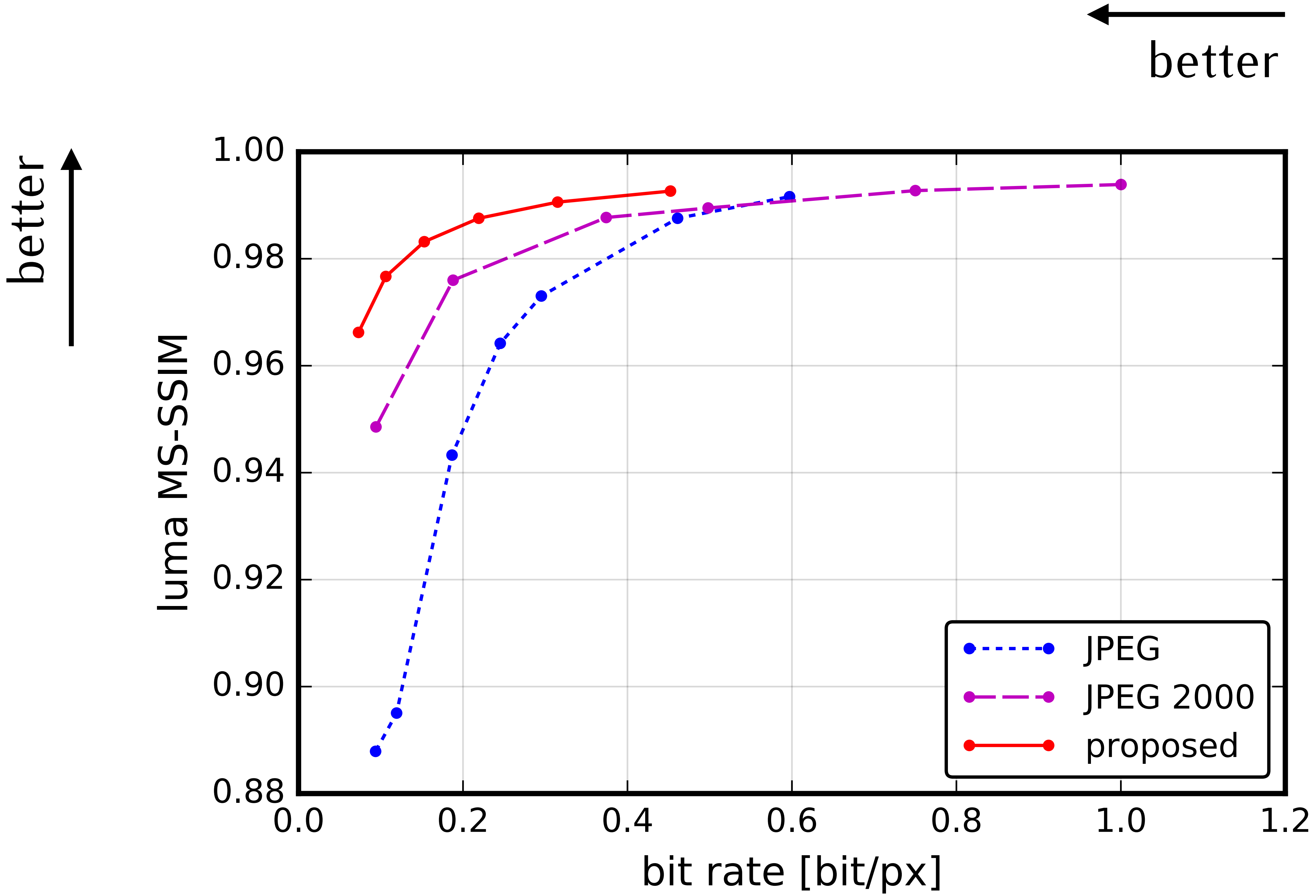
proposed @ 0.106 bits/px

original proposed



JPEG JPEG 2000

We consistently outperform JPEG 2000



original

JPEG @ 0.170 bits/px

JPEG 2000 @ 0.167 bits/px

proposed @ 0.167 bits/px

original proposed

JPEG JPEG 2000

Thanks!

More images, metrics, and the model parameters:
http://www.cns.nyu.edu/~lcv/iclr2017/

Comparison to compression state-of-the-art (BPG):
come to our poster tomorrow morning!