Coding Design of Positional Information for Robust Morphogenesis

Yoshihiro Morishita[†‡] and Yoh Iwasa[†]

[†]Department of Biology, Faculty of Sciences, Kyushu University, Fukuoka, Japan; and [‡]PRESTO, Japan Science and Technology Agency, Saitama, Japan

Supporting Materials

A. Calculation of reproducibility of differentiation shown in Fig.1

As explained in "Model" section of the main text, the spatial profiles of morphogens determine the way the spatial coordinates in real space are mapped into the concentration coordinates in chemical space (i.e., the encoding rule). On the other hand, cells read out or estimate their spatial coordinates ($\hat{\boldsymbol{x}}$) from the observed set of morphogen concentrations ($\boldsymbol{u}'$). The way of readout gives a decoding rule. We assume the maximum likelihood decoding, $\hat{\boldsymbol{x}}_{ML}(\boldsymbol{u}')$. In the presence of variability in the spatial profiles of morphogens, the estimated position $\hat{\boldsymbol{x}}_{ML}(\boldsymbol{u}')$ may in general be different from the true position $\boldsymbol{x}$.

In Fig. 1, each morphogen has an exponential gradient of its concentration from its source along a line, S_i. The direction of S_i in real space is $(0,1)$, $(\cos 120^\circ, \sin 120^\circ)$, and $(\cos 120^\circ, -\sin 120^\circ)$ in Fig. 1(A), while, $(0,1)$, $(\cos 160^\circ, \sin 160^\circ)$, and $(\cos 160^\circ, -\sin 160^\circ)$ in Fig. 1(B). The gradient of morphogen 1 is given by $u_1(x_1, x_2) = \exp(x_1 - 2)$. As a noise source, we adopted the variability of morphogen source levels, where the noise for each morphogen was assumed to be independent of each other and the standard deviation was set to $\sigma_i = 0.1$.

Since the noise level is not so large, the maximum-likelihood estimate of position $\hat{\boldsymbol{x}}_{ML}(\boldsymbol{u}')$ for observed concentrations $\boldsymbol{u}'$ is approximately given by:

$$\hat{\boldsymbol{x}}_{ML} = \boldsymbol{x} + \left[D^T \Sigma^{-1} D\right]^{-1} D^T \Sigma^{-1} \left(\boldsymbol{u}' - \boldsymbol{u}(\boldsymbol{x})\right). \qquad \text{[A1]}$$

We assume that cells differentiate if $\hat{\boldsymbol{x}}_{ML}(\boldsymbol{u}')$ is included in the target shape T in real space and do not differentiate otherwise. Then, the reproducibility of differentiation at each position $\boldsymbol{x}$ is defined as the probability $P_{dif}(\boldsymbol{x})$ that $\hat{\boldsymbol{x}}_{ML}(\boldsymbol{u}')$ is included in the shape T, which is equivalent to the success probability of differentiation. $P_{dif}(\boldsymbol{x}) = 1$ always holds in the absence of noises. The right panels of Fig. 1 show the density plots of $P_{dif}(\boldsymbol{x})$.

B. Some assumptions in the mathematical formulation of positional information coding

In this study, we considered a static situation in which the tissue growth and the change of average source levels of morphogens are negligible. In some cases, this assumption is biologically plausible since their time scales are much slower than those

for chemical reactions and diffusion. Under the assumptions, the average profiles of morphogen concentrations are time-invariant (i.e., $\boldsymbol{u}(\boldsymbol{x},t)=\boldsymbol{u}(\boldsymbol{x})$), and $\boldsymbol{u}(\boldsymbol{x})$ is defined as the encoding rule of positional information. The noise dealt with in this study is assumed to include all of the followings: the variability in source levels, source locations, diffusivity, and embryo size. Their averages are regarded as time-invariant. The variability causes the uncertainty of the concentrations observed by cells, which defines the probability distribution $P(\boldsymbol{u}';\boldsymbol{u}(\boldsymbol{x}))$.

Although there may exist temporal fluctuations in the spatial profiles of morphogen concentrations, in the dynamics of ligand-receptor binding, or during an intracellular signaling cascade, we think that some parts of them are cancelled by time-averaging (15) and/or negative feedback regulations (Becskei and Serrano (2000) *Nature* 405:590–593; Morishita et al. (2005) *J. Theor. Biol.*, 235: 241-264.). For decoding, we assume that cells are clever enough so that they can do maximum likelihood estimation of their position from the observed morphogen concentrations, and we do not ask "to what extent cells can achieve the ML-decoding". Deriving the optimal encoding under the assumptions of static variability and ML-decoding is still meaningful because it can be a criterion to evaluate the optimality of the coding in analyzing experimental data (see also "Summary and discussion" and Fig. 7 for a possible procedure of data analysis).

We assumed that the weight (or penalty) of the error in the positional estimate is spatially-isotropic; in other words, the deviation of position along an axis (e.g., the proximo-distal axis) and that along another axis (e.g., the antero-posterior axis) are equally harmful to normal development. Even if the weight of importance differs between the axes, we would obtain the same results by spatial rescaling based on appropriate weights. We can even adopt measures of precision different from the one adopted in this paper, namely, the inverse of the determinant of the variance-covariance matrix for the estimated position (i.e., $1/\det[\mathrm{Var}(\hat{\boldsymbol{x}})]$). However, for general measures of precision, no inequality, such as Cramer-Rao's inequality [1], for evaluating the order of goodness of estimation of position is available, making general arguments impossible. Even in such situations, however, the general results regarding the best way of encoding obtained in this study can still be useful as a standard reference.

For practical purposes, in inequality [1], we assume that $\mathrm{Var}[\hat{\boldsymbol{x}}]$ is a positive definite matrix and that $\hat{\boldsymbol{x}}$ is an unbiased estimator.

C. Geometrical interpretation of $\det[I(x)]$ maximization

Suppose that two cells located close to each other, say at $\boldsymbol{x}$ and $\boldsymbol{x}+\Delta\boldsymbol{x}$, detect morphogen concentrations whose values obey the probability density functions $P(\boldsymbol{u}';\boldsymbol{u}(\boldsymbol{x}))$ and $P(\boldsymbol{u}';\boldsymbol{u}(\boldsymbol{x}+\Delta\boldsymbol{x}))$ (see Fig. S1(A)). If the two distributions overlap only slightly, the probability that the two cells detect different morphogen concentrations ($\boldsymbol{u}'$) is higher, which makes distinguishing their positions on the basis of

$\boldsymbol{u}'$ easy. In contrast, if the overlap is large, the two cells are likely to detect similar concentrations, making their positions difficult to distinguish on the basis of $\boldsymbol{u}'$. In information theory, Kullback-Leibler divergence D_{KL} is often adopted as a natural measure of the amount of overlap of two probability density functions $Q_1(\boldsymbol{u})$ and $Q_2(\boldsymbol{u})$. D_{KL} is defined as follows (Ref. (23)):

$$D_{KL}(Q_1(\boldsymbol{u}), Q_2(\boldsymbol{u})) \equiv \int Q_1(\boldsymbol{u}) \ln \frac{Q_1(\boldsymbol{u})}{Q_2(\boldsymbol{u})} \mathrm{d}\boldsymbol{u}. \qquad \text{[C1]}$$

When $\Delta\boldsymbol{x}$ is small, D_{KL} for $P(\boldsymbol{u}'; \boldsymbol{u}(\boldsymbol{x}))$ and $P(\boldsymbol{u}'; \boldsymbol{u}(\boldsymbol{x}+\Delta\boldsymbol{x}))$ approximately becomes a quadratic form including Fisher information matrix as follows:

$$\begin{aligned}
&D_{KL}(P(\boldsymbol{u}', \boldsymbol{x}), P(\boldsymbol{u}', \boldsymbol{x}+\Delta\boldsymbol{x})). \\
&\equiv \int P(\boldsymbol{u}', \boldsymbol{x}) \ln \frac{P(\boldsymbol{u}', \boldsymbol{x})}{P(\boldsymbol{u}', \boldsymbol{x}+\Delta\boldsymbol{x})} \mathrm{d}\boldsymbol{u}' \\
&\cong \frac{1}{2} \int P(\boldsymbol{u}', \boldsymbol{x}) \left(\sum_{i,j} \frac{\partial \ln P}{\partial x_i} \frac{\partial \ln P}{\partial x_j} \Delta x_i \Delta x_j \right) \mathrm{d}\boldsymbol{u}' \\
&= \frac{1}{2} \sum_{i,j} I_{ij}(\boldsymbol{x}) \Delta x_i \Delta x_j \\
&= \frac{1}{2} \Delta\boldsymbol{x}^T I(\boldsymbol{x}) \Delta\boldsymbol{x} \qquad \text{[C2]}
\end{aligned}$$

In this manner, the Fisher information matrix determines the overlap (or the distance) between two probability distributions. This distance can be also interpreted as the distance between two cells measured in chemical space. It should be noted that Kullback-Leibler divergence is a generally asymmetric quantity with respect to probability distributions $Q_1(\boldsymbol{u})$ and $Q_2(\boldsymbol{u})$. However, for small $\|\Delta\boldsymbol{x}\|$, it can be approximately regarded as symmetrical,

$$D_{KL}(P(\boldsymbol{u}', \boldsymbol{x}), P(\boldsymbol{u}', \boldsymbol{x}+\Delta\boldsymbol{x})) \cong D_{KL}(P(\boldsymbol{u}', \boldsymbol{x}+\Delta\boldsymbol{x}), P(\boldsymbol{u}', \boldsymbol{x})).$$

Figure S1(B) shows examples of the value of D_{KL} around the true position $\boldsymbol{x}$ for 1D and 2D positioning. Maximizing $\det[I(\boldsymbol{x})]$ (i.e., the precision of the positional information) is equivalent to minimizing the size of the region occupied by $\Delta\boldsymbol{x}$ that satisfies $D_{KL} \leq \delta$ for a given δ (red regions). The value of δ can be interpreted as the limit of distinction in the chemical space below which the position $\boldsymbol{x}$ and $\boldsymbol{x}+\Delta\boldsymbol{x}$ are indistinguishable on the basis of observed morphogen concentrations.

Figure S1

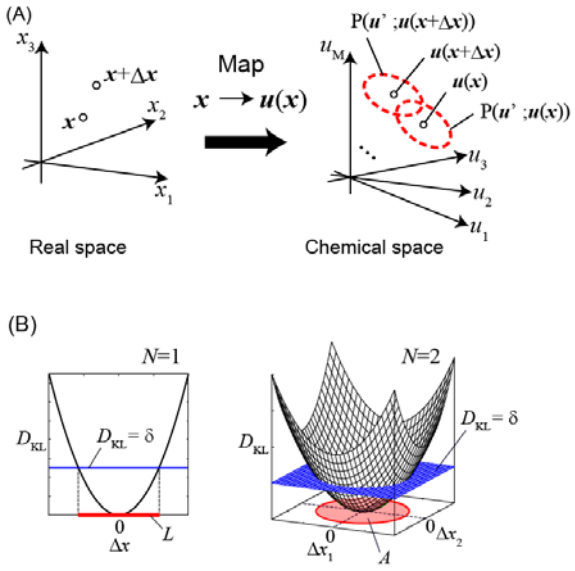

Figure legend for Fig. S1

(a) Cells located at $\boldsymbol{x}$ and $\boldsymbol{x}+\Delta\boldsymbol{x}$ detect morphogen concentrations whose values obey the probability density functions $P(\boldsymbol{u}';\boldsymbol{u}(\boldsymbol{x}))$ and $P(\boldsymbol{u}';\boldsymbol{u}(\boldsymbol{x}+\Delta\boldsymbol{x}))$.

(b) Maximizing $\det[I(\boldsymbol{x})]$ (i.e., the precision of the positional information) is equivalent to minimizing the size of the region occupied by $\Delta\boldsymbol{x}$ that satisfies $D_{KL} \le \delta$ for a given δ (red regions). The left panel shows the case of 1D positioning, and the right 2D positioning.

D: Spatial dependence of positioning performance ($\|\boldsymbol{\eta}_i\|$) and correlation coefficients (ρ_{ij}) of noises associated with morphogens.

As stated in the main text, in 1D positioning, there is experimental evidence that a morphogen gradient has constant PP ($|\eta| = const.$) within a certain spatial range (4, 19). For example, the constant PP holds for an exponential gradient if the noise is proportional to its average concentration ($\sigma_i(x) \propto u_i(x)$) at each x. When the main noise source is the variability of morphogen source levels, such a situation can be achieved.

On the other hand, in the multi-dimensional case, there is no experimental evidence supporting that PP can be regarded as being independent of position. However, in our previous study, we numerically observed that, in the context of vertebrate limb bud development, PP in the 2D case is almost constant over a wide area when morphogen gradients are formed by simple diffusion and linear degradation (leading to a nearly exponential gradient in a focal region) and when the main noise source is the variability of morphogen source levels obeying Gaussian distribution (18).

As for correlation coefficients, in a previous study (19), we showed that the variability of concentrations of Bicoid and Caudal has positive correlation at each x and its value is almost constant in a wide range of an embryo.

E. Relation between the precision of positional information ($\det[I(x)]$) and the number of morphogens (M) when different morphogens have the same PP and their noises are independent of each other.

When different morphogens have the same PP (i.e., $|\boldsymbol{\eta}_i| = \eta$ for all i) and their noises are independent of each other ($\rho_{ij} = 0$ for all (i, j)), the precision $\det[I(\boldsymbol{x})]$ is proportional to $(M/N)^N$, where M and N are the number of morphogen species

and the number of spatial dimensions, respectively. In 1D positioning, $\det[I(x)] = I(x) = \sum_i \eta_i^{\,2} = M\eta^2$ holds (see the black dotted line in Fig. S2). In the 2D (or 3D) case, the precision, $\det[I(\boldsymbol{x})]$, is the sum of the squared areas (or volumes) spanned by all possible pairs of PP vectors, i.e., $\det[I(\boldsymbol{x})] = \sum_{i<j} \left|\boldsymbol{\eta}_i \times \boldsymbol{\eta}_j\right|^2$ (or $\det[I(\boldsymbol{x})] = \sum_{i<j<k} \left|\boldsymbol{\eta}_i \cdot \left(\boldsymbol{\eta}_j \times \boldsymbol{\eta}_k\right)\right|^2$). The red and blue points in Fig. S2 are the maximum values found by numerical simulations where the value of $\det[I(\boldsymbol{x})]$ was calculated for different combinations of directions of PP vector $\boldsymbol{\eta}_i$. Those numerically searched maximums are consistent with the curve $(M/N)^N \eta^{2N}$.

Figure S2

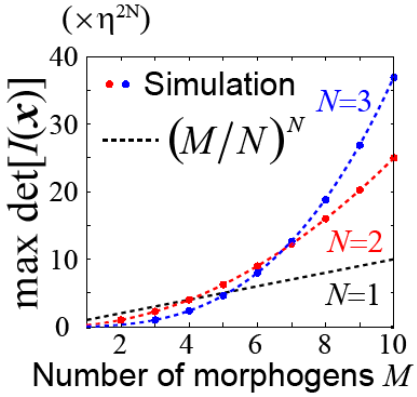

Figure legend for Fig. S2
Relation between the precision of positional information and the number of morphogens. See Supporting Materials E for details.

F. 1D positioning with multiple morphogens

In general, the impact of relative gradient directions of a pair of morphogens (i, j) on the precision of the positional information is given by $\left|\zeta_{ij}\right| \equiv \left|\tilde{C}_{ij}\eta_i\eta_j\right|$, where $\tilde{C}_{ij}$ is the (i, j)-cofactor of the correlation coefficient matrix C.

As the number of morphogens M increases, the conditions determining the best arrangement of morphogen sources gets more complex. However, in the specific case in which the morphogens can be decomposed into some independent groups in which the correlation of noises is negligibly small between groups and significant within groups, we can efficiently find the best configuration. Let us consider m_0 groups with independent noises. Then, the precision of the positional information can be decomposed as the sum of the precision provided by each uncorrelated m_0 group of morphogens ($I_j(x)$) as follows:

$$I(x) = D^T \Sigma^{-1} D = \sum_{j=1}^{m_0} D_j{}^T \Sigma_j{}^{-1} D_j = \sum_{j=1}^{m_0} I_j(x),$$

$$\Sigma = \begin{pmatrix} \Sigma_1 & & & & \\ & \Sigma_2 & & O & \\ & & \ddots & & \\ & O & & \Sigma_{m_0-1} & \\ & & & & \Sigma_{m_0} \end{pmatrix}. \qquad \text{[F1]}$$

Therefore, the best placement of morphogen sources can be independently determined for each group.

G. Relation between information entropy and Fisher information

In the absence of informational redundancy (i.e., $M = N$), $P(\hat{\boldsymbol{x}})$ obeys a Gaussian distribution with $Var(\hat{\boldsymbol{x}}) = (D^T \Sigma^{-1} D)^{-1}$ for Gaussian noise Σ, and thus the information entropy for $P(\hat{\boldsymbol{x}})$ is given by:

$$H(P(\hat{\boldsymbol{x}})) = \frac{1}{2}\log_2\left(\det[Var(\hat{\boldsymbol{x}}_{\boldsymbol{ML}})]\right) + \text{constant} = -\frac{1}{2}\log_2\left(\det[I(\boldsymbol{x})]\right) + \text{constant}\,. \quad \text{[G1]}$$

Therefore, maximizing $\det[I(\boldsymbol{x})]$ is equivalent to minimizing the entropy. This result supports that the mathematical formulation developed in this study is a natural extension of the one developed previously (18).

H. Proof that symmetrical crossing is the best encoding when different morphogens have the same PP and their noises are independent.

When all morphogens have the same PP (i.e., $|\boldsymbol{\eta}_i| = \eta = \text{const.}$) and have uncorrelated noises, the best encoding to maximize the precision is achieved when the directions of morphogen gradients are arranged symmetrically: $\theta_i = 2\pi(i-1)/M + q_i\pi$, where θ_i is the angle between the gradient vector of morphogen i and the x-axis, and q_i is an arbitrary integer. This follows from $\partial \det[I]/\partial\theta_i = 0$ for every i ($i = 1, \cdots, M$):

$$\begin{aligned} \frac{\partial \det[I]}{\partial \theta_k} &\propto \sum_{i\neq k} \frac{\partial}{\partial \theta_k} \sin^2(\theta_k - \theta_i) \\ &= \sum_{i\neq k} \sin\left(2(\theta_k - \theta_i)\right) \\ &= \sum_{i\neq k} \sin\frac{4\pi(k-i)}{M} = 0\,. \end{aligned} \qquad \text{[H1]}$$

We also confirmed by numerical simulations that the symmetrical crossing is the best encoding under the above conditions.

I. The upper limit of the precision of positional information increases with the magnitude of the correlation.

As shown in Fig. S3, when the magnitude of correlation between the variability of morphogen concentrations is larger, the maximum value of $\det[I(\boldsymbol{x})]$ increases. The

figure shows the case of 2D positioning by 3 morphogens under the conditions that $\|\boldsymbol{\eta}_1\| = \|\boldsymbol{\eta}_2\| = \|\boldsymbol{\eta}_3\|$, $\rho_{12} = \rho_{31} = 0$, and $\rho_{23} \neq 0$. $\max \det[I(\boldsymbol{x})]$ was found by numerically calculating $\det[I(\boldsymbol{x})]$ for different sets of $(\theta_{12}, \theta_{31})$, where PP vectors for 3 morphogens are given by $\boldsymbol{\eta}_1 = (1,0)$, $\boldsymbol{\eta}_2 = (\cos\theta_{12}, \sin\theta_{12})$, $\boldsymbol{\eta}_3 = (\cos(-\theta_{31}), \sin(-\theta_{31}))$, respectively. This result indicates that the precision of positional information can be improved by increasing the correlation of noises, as well as by increasing the informational redundancy (see Supporting Materials E).

Figure S3

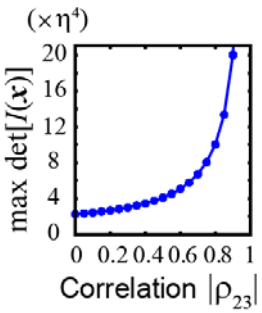

Figure legend for Fig. S3
Relation between the maximum precision of positional information and correlation of noises between morphogens. See Supporting Materials I for details.

J. A possible procedure of data analysis to validate coding optimality
Here we explain the details of the procedure to validate coding optimality shown in Fig. 7.

(STEP 1) For given quantitative data about spatial profiles of morphogen concentrations for different embryos ($\boldsymbol{u}^{(i)}(\boldsymbol{x})$), PP vectors can be calculated from the statistics, average profile ($\boldsymbol{u}(\boldsymbol{x})$) and variability of spatial profiles ($\Sigma(\boldsymbol{x})$). The first knowledge from the data analysis is PP ($\|\boldsymbol{\eta}_i(\boldsymbol{x})\|$) for each morphogen, i.e., the precision of positional information along the direction of $\boldsymbol{\eta}_i(\boldsymbol{x})$ provided by the morphogen.

(STEP 2) From $\boldsymbol{u}(\boldsymbol{x})$ and $\Sigma(\boldsymbol{x})$, the local precision of positional information ($\det[I(\boldsymbol{x})]$) for the given encoding $\boldsymbol{u}(\boldsymbol{x})$ can be calculated. $\det[I(\boldsymbol{x})]$ is the theoretically maximum precision of positional information provided to a cell located at $\boldsymbol{x}$ under the assumption of ML decoding. By examining the spatial dependences of $\det[I(\boldsymbol{x})]$, we could predict the region with structural importance where higher precision is needed.

(STEP 3) By comparing the set of directions of morphogen gradient calculated by data

$D(\boldsymbol{x}) = \left(\boldsymbol{D}_1(\boldsymbol{x}), \cdots, \boldsymbol{D}_{\boldsymbol{M}}(\boldsymbol{x})\right)$ and the set $D^{\mathrm{OPT}}(\boldsymbol{x}) = \left(\boldsymbol{D}_1^{\mathrm{OPT}}(\boldsymbol{x}), \cdots, \boldsymbol{D}_{\mathrm{M}}^{\mathrm{OPT}}(\boldsymbol{x})\right)$ which maximizes $\det[I(\boldsymbol{x})]$ under a given constraint for $\left\|\boldsymbol{D}_i(\boldsymbol{x})\right\| = \left\|\boldsymbol{D}_{\boldsymbol{i}}^{\mathrm{OPT}}(\boldsymbol{x})\right\|$ and $\Sigma(\boldsymbol{x})$, we can evaluate the goodness of encoding at each location $\boldsymbol{x}$, where $\boldsymbol{D}_{\mathrm{i}}^{\mathrm{OPT}}(\boldsymbol{x})$ is defined as:

$$\boldsymbol{D}_{\mathrm{i}}^{\mathrm{OPT}}(\boldsymbol{x}) = \mathrm{grad}\, u_i^{\mathrm{OPT}}(\boldsymbol{x}) = \left\|\boldsymbol{D}_{\boldsymbol{i}}(\boldsymbol{x})\right\| \boldsymbol{e}_{\mathrm{i}}^{\mathrm{OPT}}(\boldsymbol{x}), \quad \left\|\boldsymbol{e}_{\mathrm{i}}^{\mathrm{OPT}}(\boldsymbol{x})\right\| = 1 \qquad \text{[J1]}$$

In addition, we will be able to evaluate the global optimality of encoding by calculating the spatial patterns of $\det[I(\boldsymbol{x})]$ for different combinations of spatial arrangements of morphogen sources as performed in Fig. 6(B). On the other hand, by examining the consistency between the directions of $\mathbf{T}_1(\boldsymbol{x})$ and $\mathbf{T}_2(\boldsymbol{x})$, we can evaluate how much the adopted decoding is close to the ML decoding, where $\mathbf{T}_1(\boldsymbol{x})$ and $\mathbf{T}_2(\boldsymbol{x})$ are the direction of the contour in the chemical space satisfying $\hat{\boldsymbol{x}}_{\boldsymbol{ML}}(\boldsymbol{u}') = \boldsymbol{x}$ and that of the contour of output levels $dF(\boldsymbol{u}(\boldsymbol{x})) = 0$ (see “Model” and Fig. 3).