Bayesian gates for reliable logical operations under noisy conditions

Tetsuya J. Kobayashi

Phys. Rev. E **101**, 042205 — Published 13 April 2020

DOI: 10.1103/PhysRevE.101.042205

Bayesian Gates for Reliable Logical Operations under Noisy Condition

Tetsuya J. Kobayashi*
Institute of Industrial Science, the University of Tokyo,
4-6-1 Komaba Meguro-ku, Tokyo 153-8505, Japan.
(Dated: March 13, 2020)

The reliability of logical operations is indispensable for the reliable operation of computational systems. Since the down-sizing of micro-fabrication generates non-negligible noise in these systems, a new approach for designing noise-immune gates is required. In this paper, we demonstrate that noise-immune gates can be designed by combining Bayesian inference theory with the idea of computation over noisy channels. To reveal their practical advantages, the performance of these gates is evaluated in comparison with a stochastic resonance-based gate proposed previously. We also demonstrate that, in a high noise-level situation, this approach for computation can be better than a conventional one that conducts information transmission and computation separately.

PACS numbers: Valid PACS appear here

Reliability of logical operations is an indispensable prerequisite for the operation of almost all computations. Because of the current demand for down-sizing and low energy consumption of computational devices, new designs of logical operations with higher noise-immunity are required. To address this problem, biological systems are regarded as good role models, because our body and brain can conduct certain computations robustly and stably, even though their elementary processes, i.e., intracellular reactions and single-neuron spikes, consume low amounts of energy and thereby operate stochastically [1, 2].

Historically, biology has in fact inspired new designs of noise immune systems. Schmitt trigger is an early example, where the study of squid nerves directly led to the idea to use hysteresis for noise-immunity [3]. More recently, a new implementation of noise-immune logical operations was proposed based on stochastic resonance (SR) [4–6], which has been observed in neural sensory systems to amplify signals with the aid of noise. Nonetheless, neither hysteresis nor SR is suffices to explain all the noise-immune properties of biological systems.

A new possible mechanism in the current spotlight is an exploitation of Bayesian computation. Recent psychological and molecular-biological studies indicated that biological systems ubiquitously employ Bayesian logic for computations under noise and uncertainty from cognitive down to molecular level [7–9]. This fact suggests that the employment of the Bayesian logic for computation can contribute not only to high-level algorithms but also to low-level gate-design.

In this paper, we demonstrate that Bayesian logic can in fact be utilized for designing noise-immune logic gates in combination with the idea of computation over noisy channels presented in [10]. Bayesian logic gates are shown to have several advantageous properties over the gates based on the previously proposed logical SR (LSR). In addition, when the noise-level is sufficiently high, a scheme that operates computation over noisy channels (Fig. 1 (A)) can be better than a conventional one where information transmission and computation are conducted separately (Fig. 1 (B)). Finally, the generality and possible extensions of this approach are discussed.

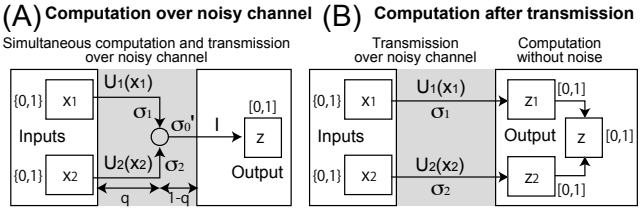

FIG. 1: Schematic diagram of Bayesian logic gate (A) and the conventional noiseless computation after information transmission (B).

Let $x_1(t) \in \{0,1\}$ and $x_2(t) \in \{0,1\}$ be two noiseless binary inputs to a logic gate at time t. The binary input $\boldsymbol{x}(t) = (x_1(t), x_2(t))^{\mathbb{T}} \in \{0,1\}^2$ is generally implemented by a physical state such as voltage as $U_i(t) = \mu_i(\boldsymbol{x}(t)) \in \mathbb{R}$ for $i \in \{1,2\}$. If noise in the physical inputs $\boldsymbol{U}(t) := (U^1(t), U^2(t))^{\mathbb{T}} \in \mathbb{R}^2$ is negligible, a logical operation over $\boldsymbol{x}(t)$, e.g., AND operation $x_1(t) \bigwedge x_2(t)$, can be implemented by a two-state thresholded dynamics in which the state flips only when both $U_1(t)$ and $U_2(t)$ exceed certain threshold values. However, when the noise in $\boldsymbol{U}(t)$ is sufficiently strong, such dynamics leads to erroneous switching driven by the noise.

The influence of noise in $U_i(t)$ is abstractly modeled in this work by the white Gaussian noise (Fig. 2) as

$$U_i(t)\mathrm{d}t = \mu_i(\boldsymbol{x}(t))\mathrm{d}t + \sigma_i \mathrm{d}W_t^i, \qquad i \in \{1,2\}.$$

Here, $\sigma_i W_t^i$ is the one-dimensional Wiener process that represents noise with intensity $\sigma_i > 0$. We also assume that $U_i(t)$ depends only on $x_i(t)$, i.e., $\mu_i(\boldsymbol{x}(t)) =$

*Electronic address: tetsuya@mail.crmind.net; URL: http://research.crmind.net/

$\mu_i(x_i(t))$. When the signal-to-noise ratio (SNR) $\Delta\mu_i/\sigma_i$ is not sufficiently high, where $\Delta\mu_i := |\mu_i(1) - \mu_i(0)|$, the simple threshold-based switching fails to return the correct output of, for example, the AND operation, because the noisy $U_1(t)$ and $U_2(t)$ can exceed the thresholds, even when $x_1(t) = 1$ and $x_2(t) = 1$ do not hold. This fact illustrates that a simple threshold-based switching does not suffice to implement a reliable logical operation under noise. To overcome this problem, we need a dynamical implementation of logical operations that is more reliable than the simple switching dynamics.

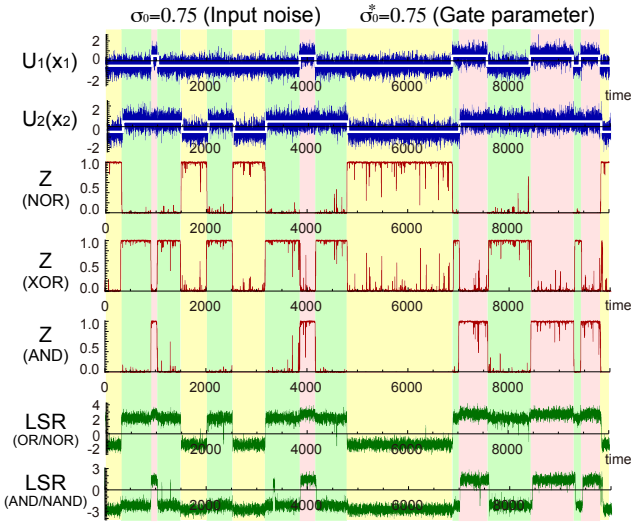

FIG. 2: (A) Sample paths of the noiseless inputs $x_1(t)$ and $x_2(t)$ (white lines), the noisy inputs $U_1(t)$ and $U_2(t)$(blue curves), the outputs $z(t)$ of the Bayesian NOR, XOR and AND gates (red curves), and the outputs $y(t)$ (green curves) of the LSR OR/NOR and AND/NAND gates. The noise amplitude of $U_1(t)$ and $U_2(t)$ is appropriately scaled for visibility. The time intervals within which the errorless gate outputs should be 1 are represented by the filled yellow (NOR), green (XOR), and red (AND) regions. Parameters are $\mu = 1$, $\sigma_0 = \sigma_0^* = 0.75$, $r_{\rm on} = r_{\rm off} = 1/1000$, $y_l^* = -0.5$, $y_u^* = 1.3$, $\alpha = 1.8$, and $\beta = 3$.

To theoretically derive such an implementation, in this work, we reformulate logical operations under noise as a statistical inference of partial information. In the conventional statistical inference and logic operations, we infer all hidden states of $\boldsymbol{x}(t)$ from $\boldsymbol{U}(t)$ as $\boldsymbol{z}(t) = (\boldsymbol{z}_1(t), \boldsymbol{z}_2(t))^{\mathbb{T}}$, where $\boldsymbol{z}(t)$ is the inferred version of $\boldsymbol{x}(t)$. After inference of transmitted information, the functions of $\boldsymbol{x}(t)$ are calculated with $\boldsymbol{z}(t)$ under noiseless conditions (Fig. 1 (B)). However, the approach here differs from the usual statistical inference in that the main purpose is the inference of partial information of $\boldsymbol{x}(t)$, i.e., a function of $\boldsymbol{x}(t)$, rather than the entire information of $\boldsymbol{x}(t)$, because a logical operation constitutes a reduction of the information on $\boldsymbol{x}(t)$ that $\boldsymbol{U}(t)$ posseses. This property enables us to conduct the necessary computation over noisy signal $\boldsymbol{U}(t)$ before inference, as shown in Fig. 1 (A). For example, $I(t) = U_1(t) + U_2(t)$ conveys sufficient information to obtain AND, NAND, OR, NOR, and XOR operations. This operation over noisy $\boldsymbol{U}$ substantially reduces the complexity of computation after inference by avoiding the calculation of the desired output of a gate from the inferred states of $\boldsymbol{x}(t)$. Therefore, the concept of computation over noisy signals (channels) is suitable for implementing a reliable circuit by combining unreliable and reliable components effectively, and may also be relevant for biological systems.

To demonstrate this idea, in this paper, we consider only $I(t) = U_1(t) + U_2(t)$ as computation over noisy $\boldsymbol{U}(t)$, although this approach is applicable for more general situations. From the definition of $U_i(t)$ and the properties of the Wiener process, $I(t)$ can be simplified as $I(t)\mathrm{d}t = \nu(\boldsymbol{x}_t)\mathrm{d}t + \sigma_0 \mathrm{d}W_t$, where $\nu(\boldsymbol{x}(t)) = \mu_1(x_1(t)) + \mu_2(x_2(t))$. The noise intensity σ_0 depends on the physical implementation of the gate. For the worst case, where U_1 and U_2 add up just before the inference computation, the noise both in U_1 and U_2 contributes to σ_0 as $\sigma_0^2 = \sigma_1^2 + \sigma_2^2$. In contrast, for the best case, the noise of the single channel that transmits $I(t)$ contributes to the noise as $\sigma_0^2 = {\sigma'}_0^2$. In order to account for these two extreme situations, we introduce a parameter $q \in [0, 1]$ such that $\sigma_0^2 = q(\sigma_1^2 + \sigma_2^2) + (1 - q){\sigma'}_0^2$ as in Fig. 1 (A) and, for simplicity, assume that $\sigma_1 = \sigma_2 = {\sigma'}_0 = \sigma$ to obtain $\sigma_0^2 = (1 + q)\sigma^2$. In addition, we also assume that $\mu_1(x) = \mu_2(x) = \mu(x)$ because of the symmetry of logic gates with respect to the exchange of two inputs. Thus, for sufficiently small $\Delta t > 0$ and fixed $\boldsymbol{x}(t)$, the probability distribution for $I(t)$ can be represented as $\mathbb{P}_N(I(t); \nu(\boldsymbol{x}(t)), \sigma_0^2/\sqrt{\Delta t})$, where $\mathbb{P}_N(I; \nu, \sigma^2)$ is the normal distribution, the mean and variance of which are ν and σ^2, respectively. Because of its definition, $\nu(\boldsymbol{x}_t)$ is either $2\mu(1)$, $\mu(1) + \mu(0)$, or $2\mu(0)$. Thus, $I(t)$ can discriminate three of the four possible states of $\boldsymbol{x}_t$. We designate the three states by χ_i as $\chi_1 = (0, 0)$, $\chi_2 = (0, 1)$ or $(1, 0)$, and $\chi_3 = (1, 1)$. Furthermore, without losing generality, we assume that $\mu(1) = \mu/2$ and $\mu(0) = -\mu/2$. Now, the inference of $\boldsymbol{x}(t)$ from the noisy input $I(t)$ is reduced to the problem of determining whether $\boldsymbol{x}_t$ is in any of χ_i. If we infer whether $\boldsymbol{x}(t)$ is in χ_3 or not, then the inference is equivalent to the AND operation, because $\boldsymbol{x}(t) = \chi_3$ only when $\boldsymbol{x}(t) = (1, 1)$. Similarly, we can construct NOR and XOR.

The statistically optimal inference of $\boldsymbol{x}_t$ is derived by using the sequential Bayesian inference as in [8]. Let $z_i(t) := \mathbb{P}_t(\boldsymbol{x}(t) = \chi_i | I(0 : t))$ be the posterior probability that $\boldsymbol{x}(t) = \chi_i$ given the history of $I(t')$ from time $t' = 0$ to $t' = t$. By following sequential Bayes' theorem[11], we have

$$z_i(t') \propto \mathbb{P}_N\left(I(t'); \nu(\chi_i), \frac{\sigma_0^2}{\sqrt{\Delta t}}\right) \sum_j \mathbb{P}_T(t', \chi_i | t, \chi_j) z_j(t)$$

where $t' = t + \Delta t$, and $\mathbb{P}_T(t', x_i | t, x_j)$ is the transition probability that $\boldsymbol{x}(t')$ becomes χ_i when $\boldsymbol{x}(t) = \chi_j$. Then,

we have

$$\frac{z_i(t')}{z_j(t')} = \frac{\mathbb{P}_N(I(t');\nu(\chi_i),\frac{\sigma_0^2}{\sqrt{\Delta t}})}{\mathbb{P}_N(I(t');\nu(\chi_j),\frac{\sigma_0^2}{\sqrt{\Delta t}})} \frac{\sum_k \mathbb{P}_T(t',\chi_i|t,\chi_k)z_k(t)}{\sum_k \mathbb{P}_T(t',\chi_j|t,\chi_k)z_k(t)}.$$

For simplicity, we assume that $\mathbb{P}_T(t',\chi_i|t,\chi_j)$ is time-homogeneous and can be represented for sufficiently small Δt by $\mathbb{P}_T(t',\chi_i|t,\chi_j) = \Delta t \times r_{i|j}$ for $i \neq j$ and $\mathbb{P}_T(t',\chi_i|t,\chi_i) = 1 - \Delta t \times r_{i|i}$ where $r_{i|j}$ is the instantaneous transition rate from χ_j to χ_i and $r_{i|i} = \sum_{k\neq i} r_{k|i}$ holds. If the dynamics of both $x_1(t)$ and $x_2(t)$ follow a two-state Markov process whose transition rate from 0 to 1 and 1 to 0 are $r_{\rm on}$ and $r_{\rm off}$, respectively, then we have

$$\mathcal{R} := (r_{i|j}) = \begin{pmatrix} -2r_{\rm on} & r_{\rm off} & 0 \\ 2r_{\rm on} & -r_{\rm on} - r_{\rm off} & 2r_{\rm off} \\ 0 & r_{\rm on} & -2r_{\rm off} \end{pmatrix}.$$

By taking the limit as $\Delta t \to 0$, we obtain a three dimensional equation as

$$\frac{\mathrm{d}}{\mathrm{dt}}\left(\log\frac{z_i(t)}{z_j(t)}\right) = \mathcal{B}_{i,j}(I(t)) + \frac{[\mathcal{R}\boldsymbol{z}(t)]_i}{z_i(t)} - \frac{[\mathcal{R}\boldsymbol{z}(t)]_j}{z_j(t)},$$

where $\boldsymbol{z}(t) = (z_1(t), z_2(t), z_3(t))^{\mathbb{T}}$ and

$$\mathcal{B}(I(t)) = \frac{\mu}{\sigma_0^2}\left[I(t)\begin{pmatrix} 0 & -1 & -2 \\ 1 & 0 & -1 \\ 2 & 1 & 0 \end{pmatrix} + \frac{\mu}{2}\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}\right],$$

Finally, by transforming the above equation with respect to $\boldsymbol{z}(t)$, we have the following equation with quadratic nonlinearity:

$$\frac{\mathrm{d}\boldsymbol{z}(t)}{\mathrm{dt}} \overset{\circ}{=} \mathrm{diag}(\boldsymbol{z}(t))\mathcal{B}(I(t))\boldsymbol{z}(t) + \mathcal{R}\boldsymbol{z}(t), \qquad (1)$$

where $\overset{\circ}{=}$ indicates that the integrals with respect to $\mathrm{d}W_t$ are interpreted as the Stratonovich integrals [12](see Appendix for a general and detailed derivation). The three outputs, $z_1(t)$, $z_2(t)$, and $z_3(t)$, correspond to Bayesian NOR, XOR, and AND gates, respectively, and therefore, this system can simultaneously compute these operations. By an appropriate transformation of variables, we can also implement OR, NXOR, and NAND operations. Equation (1) contains σ_0 and μ as the system's parameters, indicating that the optimal tuning of these parameters such that they coincide with those of the input $I(t)$ is necessary to realize statistically optimal logical operations. However, the gate parameters may not be accurately adjusted in a real situation. In order to analyze the influence of such a parameter mismatch, we introduce μ_0^* and σ_0^* to specifically represent the gate's parameters μ and σ_0 in Eq. (1), and therefore, Eq. (1) is statistically optimal only when $\mu_0^* = \mu$ and $\sigma_0^* = \sigma_0$.

Figure 2 demonstrate that the Bayesian gates can conduct logical operations under a very noisy condition. To achieve such operations, the gates should optimally reduction noise while keeping in track of the change in the input behind noise. Additionally, the change in the input is preferably outputted in an all-or-none manner without amplifying noise. While such balancing of noise reduction and signal amplification may also be realized by optimizing a linear filter, the linear filter with the optimized response property typically requires a complex design for its implementation. In contrast, as shown above, the optimal Bayesian gates can be implemented by a relatively simple three dimensional equation with only quadratic nonlinearity and with a few control parameters. In addition, the behavior of the Bayesian gate is very robust to the increase in noise as shown in Fig. 3.

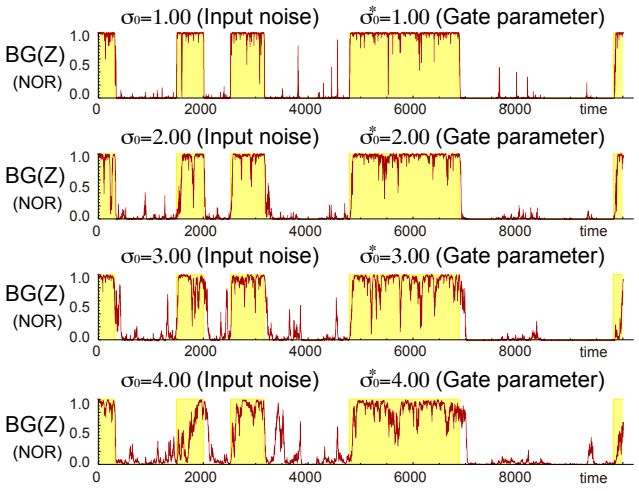

FIG. 3: Behaviors of the Bayesian NOR gate for different intensities of the input noise σ_0. The value of σ_0 is shown on each panel. The gate parameter σ_0^* is set to be optimal as $\sigma_0^* = \sigma_0$. The other parameters are the same as those in Fig.2.

We further compare the performance of the Bayesian gates with LSR, which was proposed as a nonlinear noise-immune implementation of logical gates [4, 5]. LSR is defined by the following stochastic differential equation with a double-well potential:

$$\mathrm{d}y = [-\alpha y + \beta g(y) + I(t)]\,\mathrm{dt},$$

where $g(y) = y$ when $y_l^* \leq y \leq y_u^*$, $g(y) = y_l^*$ when $y < y_l^*$, and $g(x) = y_u^*$ when $y > y_u^*$. The optimal LSR NOR gate can be implemented by setting $(y_l^*, y_u^*) = (-0.5, 1.3)$ as in [4] (Fig. 2). In order to evaluate the performance of the Bayesian and LSR NOR gates, we use a time-averaged error (TAE) defined as $\mathcal{E} := \frac{1}{T}\int_0^T |\mathbf{1}[a(t) > a_{\rm th}] - G[x_1(t); x_2(t)]|\,\mathrm{dt}$, where $a(t)$ is either $z_i(t)$ or $y(t)$. $\mathbf{1}[a > a_{\rm th}]$ returns 1 when $a > a_{\rm th}$ and 0 otherwise. For $a(t) = z_i(t)$, we choose $a_{\rm th} = 1/2$, whereas we choose $a_{\rm th} = 0$ for $a(t) = y(t)$ by following [4]. $G[x_1; x_2]$ is the output of the ideal noiseless gate with input $\boldsymbol{x}$. If the NOR gate is concerned, for example, $G[x_1; x_2]$ returns 1 when $(x_1, x_2) = (0, 0)$, and 0 other-

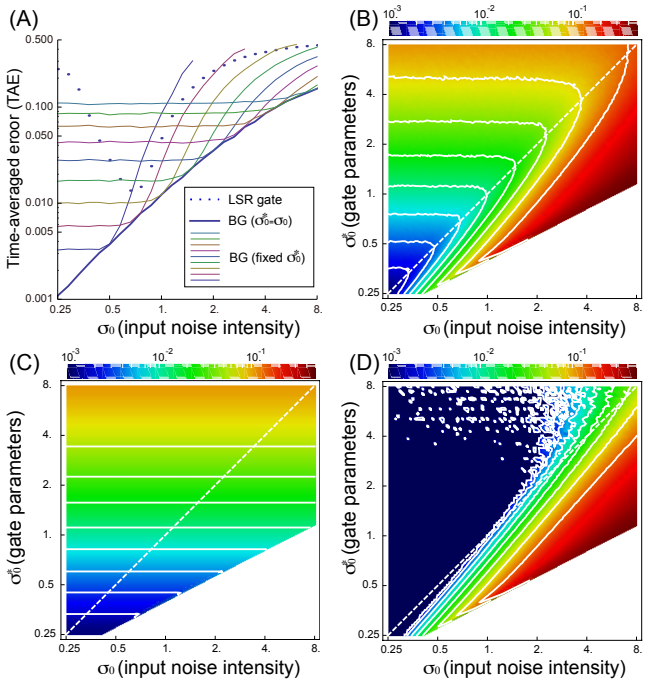

FIG. 4: (A) Time-averaged error $\mathcal{E}$ of the Bayesian NOR (solid curves) and the LSR NOR (dotted curve) gates as a function of the noise intensity σ of the input. Thin colored curves correspond to the TAE of the Bayesian NOR gate with different σ_0^*. The thick blue curve represents the TAE of the Bayesian NOR gate whose parameter σ_0^* is adjusted to be optimal as $\sigma_0^* = \sigma_0$. (B) Total TAE $\mathcal{E}$ of the Bayesian NOR gate as a function of σ_0 and σ_0^*. The broken white line represents $\sigma_0 = \sigma_0^*$. The solid white curves are the contours of TAEs. (C) Error rate of the Bayesian NOR gate due to the delay of switching, $\mathcal{E}_D$. (D) Error rate of the Bayesian NOR gate due to erroneous switching by input noise, $\mathcal{E}_E$. The other parameters than σ_0 and σ_0^* are the same as those in Fig. 2.

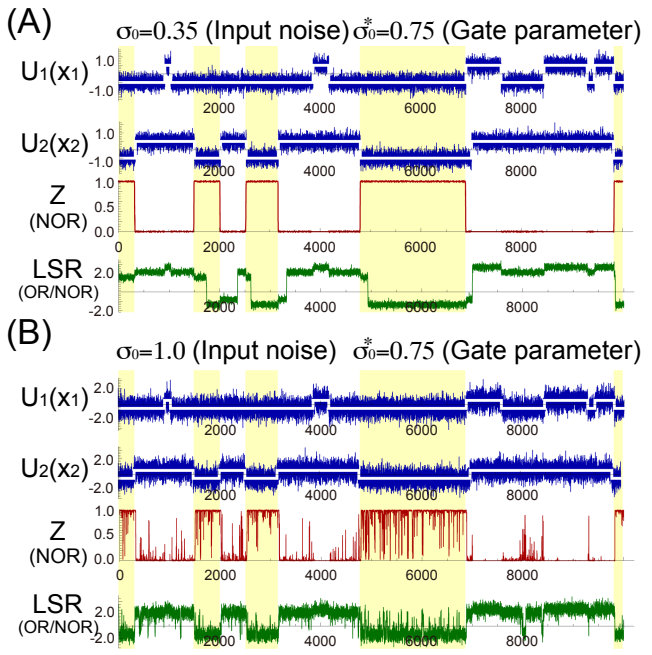

FIG. 5: Sample paths ofthe Bayesian NOR and the LSR gates when the input noise intensity σ_0 is less (A) or more (B) than the expected noise intensity σ_0^* to which the gates are optimized. All the color codes and parameters are the same as in Fig. 2 except $\sigma_0 = 0.35$ (A) and $\sigma_0 = 1.0$ (B).

wise. The performance of LSR is shown to be optimal when $0.6 < \sigma_s < 0.8$, as in Fig. 4(A). Under this optimal noise intensity for the LSR gate, the TAEs of the LSR and the Bayesian gates are comparative, indicating that the performance of the LSR gate is close to the statistical optimal (Figs. 2 and 4(A)).

As shown in Figs. 4(A) and 5, however, the performance of the LSR gate quickly degrades if the noise intensity of the input, σ_0, deviates from its optimal one, whereas the Bayesian gate can still conduct a reliable logical operation within a wider range of noise intensity (Figs. 3 and 4(B)). Furthermore, the TAE of the Bayesian gate does not increase if σ_0 is less than the gate parameter, σ_0^*. This property of the Bayesian gate means that its performance is determined by the worst noise level that σ_0^* specifies. As long as the actual noise intensity σ_0 is less than this expected worst level σ_0^*, the gate operates robustly at the cost of a fixed lower bound of the TAE. Since information of the actual noise level within a system may not always be available before designing gates, the Bayesian gate has a practical advantage over the LSR gate. In general, the error stems from either erroneous switching for constant x_t or delay in the switching of z_t when x_t changes. Since the total TAE can be attributed only to the delay of switching when no noise exists in the input, i.e., $\sigma_0 = 0$, we can approximately dissect the total TAE $\mathcal{E}$ into the errors from the delay of switching at the changes of x_t as $\mathcal{E}_D := \lim_{\sigma_0 \to 0} \mathcal{E}$, (Fig. 4(C)) and those from the erroneous switching for constant x_t as $\mathcal{E}_E := \mathcal{E} - \mathcal{E}_D$, (Fig. 4(D)). As clearly seen in Fig. 4(D), $\mathcal{E}_E$ hardly changes if σ_0^* (the expected noise intensity) is larger than σ_0 (the actual noise intensity), whereas $\mathcal{E}_D$ increases. Thus, the cost of choosing σ_0^* larger than σ_0 is the delay of switching, which limits the speed of the gate. However, σ_0^* larger than σ_0 works as a margin for systematic variations of σ_0, because $\mathcal{E}_E$ increases little provided that σ_0 is less than σ_0^*. The same result is obtained for Bayesian XOR and AND gates (Fig. 6). Thus, the variation in σ_0 can be compensated at the cost of slow switching, reflecting a trade-off between the computational speed and the reliability of computation [13].

Information transmission and computation are usually separated in conventional computational architectures in which computation is conducted under virtually noiseless conditions (Fig. 1(B)). This usual computation without noise is expected to perform better than the Bayesian gates that conducts computation and transmission simultaneously in noisy conditions. However, it requires two independent channels for input transmission (Fig. 1

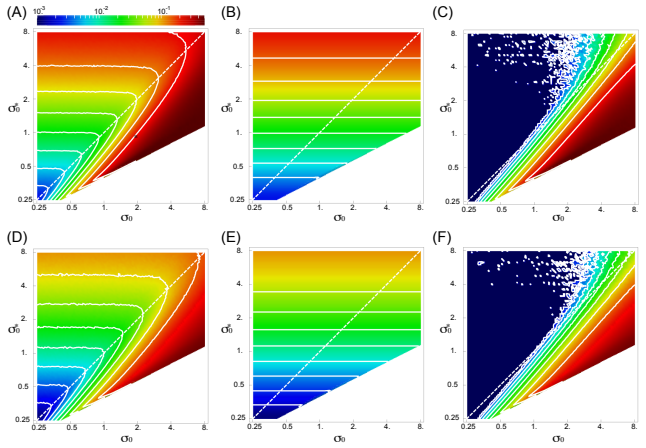

FIG. 6: The time-averaged error of the Bayesian XOR and AND gates. $\mathcal{E}$(A), $\mathcal{E}_D$(B), and $\mathcal{E}_E$(C) for Bayesian XOR gate. $\mathcal{E}$(D), $\mathcal{E}_D$(E), and $\mathcal{E}_E$(F) for Bayesian XOR gate. The broken white line and solid white curves represent $\sigma_0 = \sigma_0^*$ and the contours of error rates, respectively. All the parameters other than σ_0 and σ_0^* are the same as those in Fig. 2.

(B)), whereas the Bayesian gate combines them along the transmission path for the computation (Fig. 1 (A)). This observation suggests that the Bayesian gate may outperform the usual computation if a noisy channel is effectively exploited for computation with small q. In order to clarify this condition, we calculated $\mathcal{E}(q)$ of the Bayesian NOR gate as a function of $q \in [0, 1]$, and that of the usual computation $\mathcal{E}_U$ to obtain a performance ratio η defined as $\eta := \mathcal{E}(q)/\mathcal{E}_U$. Figure 7 (A) shows that the operation of the Bayesian gate can be more efficient ($\eta < 1$) than or comparative ($\eta < 10^{0.1} \approx 1.25$) to the usual computation when q is sufficiently small. In addition, the range of q within which the Bayesian gate operates better expands as the noise intensity σ increases. The same result is obtained for other gates (Fig. 8). This result indicates that the operation of the Bayesian gate can be efficient when the noise in the channel is large, whereas the usual computation is better when the channel noise is very small, meaning that the computation over noisy channels may be practical when the noise cannot be small.

Since our approach is based not on a specific physical implementation but on the general theory of inference, it potentially has more extensions and applications than those demonstrated here. First, we can choose arithmetic operations other than addition over noisy signals $U_1(t)$ and $U_2(t)$, which lead to different noise characteristics and gate properties. For example, subtraction may lead to more reliable gates than addition in principle by canceling out the correlated noise in U_1 and U_2. Second, we can easily design noise-immune gates with more than two inputs to conduct more complicated logical operations at the cost of the complexity of individual gates. Third, noise is not restricted to Gaussian white noise. For example, an equation similar to Eq. 1 can be derived for Poisson noise, which is more relevant to gate

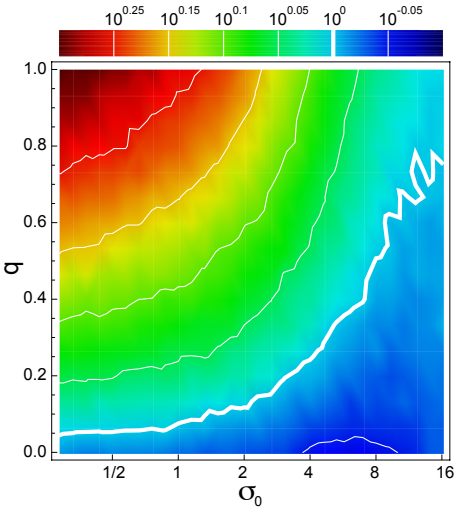

FIG. 7: The performance ratio η of the Bayesian NOR gate as a function of σ_0 and q. The white curves represent contours of η where the thick white curve corresponds to $\eta = 1$. The gate parameter is set to be optimal as $\sigma_0^* = \sigma_0$. The other parameters are the same as those in Fig. 2.

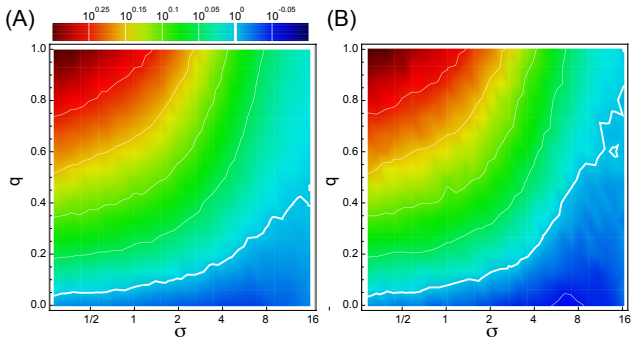

FIG. 8: Performance ratio η for the Bayesian XOR (A) and AND (B) gates. All the parameters are the same as in Fig. 7.

implementations by photons or by intracellular reactions. Since intracellular systems are known to conduct various computations [14], the Poissonian version may help us to understand and synthetically design intracellular information processing networks [15]. The combination of computation over noisy channels and the design of reliable gates by inference theory proposed in this paper is sufficiently general to cover all these situations and should be investigated further for individual situations.

We thank Hiroyasu Ando, Yuzuru Sato, Atsushi Kamimura, Yoshihiro Morishita, and Ryo Yokota for fruitful discussions. This work was supported partially by JSPS KAKENHI Grant Numbers 15H00800 and 19H05799, the Platform for Dynamic Approaches to Living Systems funded by MEXT and AMED, Japan, and the JST PRESTO program.

Appendix

A 1. DERIVATION OF BAYESIAN GATES WITH TWO INPUTS

The derivation of Bayesian gates with two gate inputs is described. In the main text, we consider the situation where two binary inputs $x_1(t)$ and $x_2(t)$ represented physically by $U_1(t)$ and $U_2(t)$ are combined into one gate input $I(t)$. In this extended derivation, we consider that $\boldsymbol{x}$ are n dimensional binary inputs, and they are combined into two gate inputs $I_1(t)$ and $I_2(t)$, which includes one gate input shown in Fig. 1(A) of the main text as a special case, as well as the conventional computation $\boldsymbol{I}(t) = \boldsymbol{U}(t)$ (Fig. 1(B) in the main text). We also consider the correlation between two inputs due to common noise as a generalization.

A. Derivation of general expression

Let $\boldsymbol{x}(t) \in \mathcal{X} := \{0,1\}^n$ be the state of an n dimensional noiseless binary input to a gate. Because of the noise and computation over a noisy channel, the actual input to the gate, $\boldsymbol{I}(t)$, is generally disturbed as

$$\boldsymbol{I}(t)\mathrm{d}t = \begin{pmatrix} \nu_1(\boldsymbol{x}(t)) \\ \nu_2(\boldsymbol{x}(t)) \end{pmatrix} \mathrm{d}t + \begin{pmatrix} \tilde{\sigma}_1 \mathrm{d}\tilde{W}_t^1 \\ \tilde{\sigma}_2 \mathrm{d}\tilde{W}_t^2 \end{pmatrix} + \tilde{\sigma}_c \begin{pmatrix} 1 \\ 1 \end{pmatrix} \mathrm{d}\tilde{W}_t^c,$$

where $I_i(t)\mathrm{d}t = \nu_i(\boldsymbol{x}(t))\mathrm{d}t + \tilde{\sigma}_i \mathrm{d}\tilde{W}_t^i + \tilde{\sigma}_c \mathrm{d}\tilde{W}_t^c$ for $i \in \{1,2\}$, and $\boldsymbol{\nu}(t) := (\nu_1(\boldsymbol{x}(t)), \nu_2(\boldsymbol{x}(t)))^{\mathbb{T}}$. We assume that $\mathrm{d}\tilde{W}_t^1$, $\mathrm{d}\tilde{W}_t^2$, and $\mathrm{d}\tilde{W}_t^c$ are independent. $\mathrm{d}\tilde{W}_t^c$ is introduced to account for the correlation due to the common noise to the gates. If we define $\sigma_i \mathrm{d}W^i = \tilde{\sigma}_i \mathrm{d}\tilde{W}_t^i + \tilde{\sigma}_c \mathrm{d}\tilde{W}_t^c$, then we have $\sigma_i^2 = \tilde{\sigma}_i^2 + \tilde{\sigma}_c^2$ for $i \in \{1,2\}$.

If more than one state of $\boldsymbol{x}(t)$ produces the same output $\boldsymbol{\nu}(\boldsymbol{x})$, the information on $\boldsymbol{x}$ is degenerated by being passed through $\boldsymbol{I}(t)$. In other words, $\boldsymbol{I}(t)$ operates as a gate to compute certain logical operations over $\boldsymbol{x}$. Let the set of subsets of $\boldsymbol{x}$ be $\mathcal{X}_I$ that can be discriminated from the output of $\boldsymbol{\nu}(\boldsymbol{x})$. Let χ_i be the i-th member of $\mathcal{X}_I$, where $i \in \{1, \ldots, \#\mathcal{X}_I\}$. From the definition, $\boldsymbol{\nu}(\boldsymbol{x}) = \boldsymbol{\nu}(\boldsymbol{x}')$ for $\boldsymbol{x}, \boldsymbol{x}' \in \chi_i$, and $\boldsymbol{\nu}(\boldsymbol{x}) \neq \boldsymbol{\nu}(\boldsymbol{x}')$ for $\boldsymbol{x} \in \chi_i$ and $\boldsymbol{x}' \in \chi_j$ with $i \neq j$. Since each χ_i has its representative member $\boldsymbol{x}_i \in \chi_i$, we identify this representative member with χ_i for notational simplicity when no confusion arises. Therefore, $\boldsymbol{x} \in \mathcal{X}_I$ is equivalent to $\chi \in \mathcal{X}_I$.

For $\boldsymbol{x} \in \mathcal{X}_I$, let $\mathbb{P}_t(\boldsymbol{x}|\boldsymbol{I}_{0:t})$ be the posterior probability of $\boldsymbol{x} \in \mathcal{X}_I$ at time t given the time series of $\boldsymbol{I}(t)$, $\boldsymbol{I}_{0:t} := \{\boldsymbol{I}(t') | t \in [0,t]\}$. By following the sequential Bayes' theorem, we have

$$\mathbb{P}_{t'}(\boldsymbol{x}'|\boldsymbol{I}_{0:t'}) \propto \mathbb{P}_{t'}(\boldsymbol{I}|\boldsymbol{x}') \sum_{\boldsymbol{x} \in \mathcal{X}_I} \mathbb{P}_T(\boldsymbol{x}', t'|\boldsymbol{x}, t) \mathbb{P}_t(\boldsymbol{x}|\boldsymbol{I}_{0:t}),$$

where $\mathbb{P}_T(\boldsymbol{x}', t'|\boldsymbol{x}, t)$ is the transition probability from $\boldsymbol{x} \in \mathcal{X}_I$ at t to $\boldsymbol{x}' \in \mathcal{X}_I$ at t', and $\mathbb{P}_{t'}(\boldsymbol{I}|\boldsymbol{x}')$ is the probability to observe $\boldsymbol{I}$ at t' conditioned by $\boldsymbol{x}'$ at t'. In general, $\mathbb{P}_T(\boldsymbol{x}' \in \mathcal{X}_I, t'|\boldsymbol{x} \in \mathcal{X}_I, t)$ may not always be well-defined, because it must be constructed from $\mathbb{P}_T(\boldsymbol{x}' \in \mathcal{X}, t'|\boldsymbol{x} \in \mathcal{X}, t)$ and the degeneration of $\mathcal{X}$ induced by $\boldsymbol{\nu}(\boldsymbol{x})$. This fact is related to the realization of a Bayesian gate, and is discussed later. At this moment, we assume that $\mathbb{P}_T(\boldsymbol{x}' \in \mathcal{X}_I, t'|\boldsymbol{x} \in \mathcal{X}_I, t)$ is well-defined, as in the case analyzed in the main text.

For $\boldsymbol{I}(t)$ defined above, $\mathbb{P}_{t'}(\boldsymbol{I}|\boldsymbol{x}')$ can be represented as

$$\mathbb{P}_{t'}(\boldsymbol{I}|\boldsymbol{x}') = \mathbb{P}_N\left(\boldsymbol{I}_{t'}; \boldsymbol{\nu}(\boldsymbol{x}'), \Sigma/\Delta t\right),$$

where $\mathbb{P}_N(\boldsymbol{I}; \boldsymbol{\nu}, \Sigma)$ is a Gaussian distribution with a mean vector $\boldsymbol{\nu}$ and covariance matrix Σ. $\boldsymbol{I}_{t'}$ here should be interpreted not as the value of $\boldsymbol{I}$ at t' but as its average between t and t'. Σ is described as

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

where the coefficient of correlation ρ is $\rho = \tilde{\sigma}_c^2/\sigma_1\sigma_2$.

For $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}_I$, we have

$$\frac{\mathbb{P}_{t'}(\boldsymbol{x}|\boldsymbol{I}_{0:t'})}{\mathbb{P}_{t'}(\boldsymbol{y}|\boldsymbol{I}_{0:t'})} = \frac{\mathbb{P}_N\left(\boldsymbol{I}_{t'}; \boldsymbol{\nu}(\boldsymbol{x}), \Sigma/\Delta t\right)}{\mathbb{P}_N\left(\boldsymbol{I}_{t'}; \boldsymbol{\nu}(\boldsymbol{y}), \Sigma/\Delta t\right)} \frac{\mathbb{P}_{t'}(\boldsymbol{x}|\boldsymbol{I}_{0:t})}{\mathbb{P}_{t'}(\boldsymbol{y}|\boldsymbol{I}_{0:t})},$$

where $\mathbb{P}_{t'}(\boldsymbol{x}|\boldsymbol{I}_{0:t}) := \sum_{\boldsymbol{x}' \in \mathcal{X}_I} \mathbb{P}(\boldsymbol{x}, t'|\boldsymbol{x}', t) \mathbb{P}_t(\boldsymbol{x}'|\boldsymbol{I}_{0:t})$. By taking the logarithm of both sides, we have

$$\begin{aligned} \log \frac{\mathbb{P}_{t'}(\boldsymbol{x}|\boldsymbol{I}_{0:t'})}{\mathbb{P}_{t'}(\boldsymbol{y}|\boldsymbol{I}_{0:t'})} =& \log \frac{\mathbb{P}_N\left(\boldsymbol{I}_{t'}; \boldsymbol{\nu}(\boldsymbol{x}), \Sigma/\Delta t\right)}{\mathbb{P}_N\left(\boldsymbol{I}_{t'}; \boldsymbol{\nu}(\boldsymbol{y}), \Sigma/\Delta t\right)} \\ &+ \log \frac{\mathbb{P}_{t'}(\boldsymbol{x}|\boldsymbol{I}_{0:t})}{\mathbb{P}_{t'}(\boldsymbol{y}|\boldsymbol{I}_{0:t})}. \end{aligned}$$

By expanding $\log \frac{\mathbb{P}_N(\boldsymbol{I}_{t'}; \boldsymbol{\nu}(\boldsymbol{x}), \Sigma/\Delta t)}{\mathbb{P}_N(\boldsymbol{I}_{t'}; \boldsymbol{\nu}(\boldsymbol{y}), \Sigma/\Delta t)}$ with respect to Δt, we have

$$\begin{aligned} \log \frac{\mathbb{P}_N\left(\boldsymbol{I}_{t'}; \boldsymbol{\nu}(\boldsymbol{x}), \Sigma/\Delta t\right)}{\mathbb{P}_N\left(\boldsymbol{I}_{t'}; \boldsymbol{\nu}(\boldsymbol{y}), \Sigma/\Delta t\right)} =& \Delta t \left[\frac{\Delta\nu_1(I_1(t') - \langle\nu\rangle_1)}{(1-\rho^2)\sigma_1^2} \right. \\ &\left. + \frac{\Delta\nu_2(I_2(t') - \langle\nu\rangle_2)}{(1-\rho^2)\sigma_2^2} - \frac{\rho(\Delta\nu_1 I_1(t') + \Delta\nu_2 I_2(t') - \Upsilon)}{(1-\rho^2)\sigma_1\sigma_2} \right] \\ &+ o(\Delta t^2), \end{aligned}$$

where

$$\begin{aligned} \Delta\nu_i = \Delta\nu_i(\boldsymbol{x}, \boldsymbol{y}) \quad &:= \nu_i(\boldsymbol{x}) - \nu_i(\boldsymbol{y}), \\ \langle\nu\rangle_i = \langle\nu\rangle_i(\boldsymbol{x}, \boldsymbol{y}) \quad &:= (\nu_i(\boldsymbol{x}) + \nu_i(\boldsymbol{y}))/2, \\ \Upsilon = \Upsilon(\boldsymbol{x}, \boldsymbol{y}) \quad &:= \nu_1(\boldsymbol{x})\nu_2(\boldsymbol{x}) - \nu_1(\boldsymbol{y})\nu_2(\boldsymbol{y}), \end{aligned}$$

and we abbreviate the dependency on $\boldsymbol{x}$ and $\boldsymbol{y}$ for readability. By redefining $\bar{\sigma}_i := \sigma_i\sqrt{1-\rho^2}$, we have

$$\begin{aligned} \log \frac{\mathbb{P}_N\left(\boldsymbol{I}_{t'}; \boldsymbol{\nu}(\boldsymbol{x}), \Sigma/\Delta t\right)}{\mathbb{P}_N\left(\boldsymbol{I}_{t'}; \boldsymbol{\nu}(\boldsymbol{y}), \Sigma/\Delta t\right)} \approx& \Delta t \left[\frac{\Delta\nu_1(I_1(t') - \langle\nu\rangle_1)}{\bar{\sigma}_1^2} \right. \\ &\left. + \frac{\Delta\nu_2(I_2(t') - \langle\nu\rangle_2)}{\bar{\sigma}_2^2} - \frac{\rho(\Delta\nu_1 I_1(t') + \Delta\nu_2 I_2(t') - \Upsilon)}{\bar{\sigma}_1\bar{\sigma}_2} \right], \end{aligned}$$

Furthermore, let us define

$$B_i(\boldsymbol{x},\boldsymbol{y}) := \frac{\Delta\nu_i(I_i(t') - \langle\nu\rangle_i)}{\bar{\sigma}_i^2},$$
$$B_i^c(\boldsymbol{x},\boldsymbol{y}) := -\frac{\rho\Delta\nu_i I_i(t')}{\bar{\sigma}_1\bar{\sigma}_2},$$
$$B_\Upsilon(\boldsymbol{x},\boldsymbol{y}) := \frac{\rho\Upsilon}{\bar{\sigma}_1\bar{\sigma}_2},$$

and then we have

$$\log\frac{\mathbb{P}_N\left(\boldsymbol{I}_{t'};\boldsymbol{\nu}(\boldsymbol{x}),\Sigma/\Delta t\right)}{\mathbb{P}_N\left(\boldsymbol{I}_{t'};\boldsymbol{\nu}(\boldsymbol{y}),\Sigma/\Delta t\right)} \approx \sum_{i\in\{1,2\}}[B_i + B_i^c] + B_\Upsilon,$$

From the definitions, we have

$$\Delta\nu_i(\boldsymbol{x},\boldsymbol{y}) = -\Delta\nu_i(\boldsymbol{y},\boldsymbol{x}),$$
$$\langle\nu\rangle_i(\boldsymbol{x},\boldsymbol{y}) = \langle\nu\rangle_i(\boldsymbol{y},\boldsymbol{x}),$$
$$\Upsilon(\boldsymbol{x},\boldsymbol{y}) = -\Upsilon(\boldsymbol{y},\boldsymbol{x}).$$

Therefore, we also have

$$B_i(\boldsymbol{x},\boldsymbol{y}) = -B_i(\boldsymbol{y},\boldsymbol{x}),$$
$$B_i^c(\boldsymbol{x},\boldsymbol{y}) = -B_i^c(\boldsymbol{y},\boldsymbol{x}),$$
$$B_\Upsilon(\boldsymbol{x},\boldsymbol{y}) = -B_\Upsilon(\boldsymbol{y},\boldsymbol{x}).$$

Furthermore, for $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\in\mathcal{X}_I$, Bs satisfies

$$B_i(\boldsymbol{x},\boldsymbol{y}) - B_i(\boldsymbol{z},\boldsymbol{y}) = B_i(\boldsymbol{x},\boldsymbol{z}),$$
$$B_i^c(\boldsymbol{x},\boldsymbol{y}) - B_i^c(\boldsymbol{z},\boldsymbol{y}) = B_i^c(\boldsymbol{x},\boldsymbol{z}),$$
$$B_\Upsilon(\boldsymbol{x},\boldsymbol{y}) - B_\Upsilon(\boldsymbol{z},\boldsymbol{y}) = B_\Upsilon(\boldsymbol{x},\boldsymbol{z}).$$

The transition probability can be expanded with respect to Δt as

$$\mathbb{P}_T(\boldsymbol{x}',t'|\boldsymbol{x},t) \approx \mathbb{I} + \Delta t\mathcal{R}(\boldsymbol{x}'|\boldsymbol{x}) + o(\Delta t^2),$$

where $\mathbb{I}$ is the identity matrix and $\mathcal{R}(\boldsymbol{x}|\boldsymbol{x}')$ is a matrix to describe the rate of transition from $\boldsymbol{x}$ to $\boldsymbol{x}'$. Thus,

$$\begin{aligned}\log\frac{\mathbb{P}_{t'}(\boldsymbol{x}|\boldsymbol{I}_{0:t})}{\mathbb{P}_{t'}(\boldsymbol{y}|\boldsymbol{I}_{0:t})} \approx& \log\frac{\mathbb{P}_t(\boldsymbol{x}|\boldsymbol{I}_{0:t}) + \Delta t\sum_{\boldsymbol{x}'}\mathcal{R}(\boldsymbol{x}|\boldsymbol{x}')\mathbb{P}_t(\boldsymbol{x}'|\boldsymbol{I}_{0:t})}{\mathbb{P}_t(\boldsymbol{y}|\boldsymbol{I}_{0:t}) + \Delta t\sum_{\boldsymbol{y}'}\mathcal{R}(\boldsymbol{y}|\boldsymbol{y}')\mathbb{P}_t(\boldsymbol{y}'|\boldsymbol{I}_{0:t})}\\ \approx& \log\frac{\mathbb{P}_t(\boldsymbol{x}|\boldsymbol{I}_{0:t})}{\mathbb{P}_t(\boldsymbol{y}|\boldsymbol{I}_{0:t})} + \Delta t\frac{\sum_{\boldsymbol{x}'}\mathcal{R}(\boldsymbol{x}|\boldsymbol{x}')\mathbb{P}_t(\boldsymbol{x}'|\boldsymbol{I}_{0:t})}{\mathbb{P}_t(\boldsymbol{x}|\boldsymbol{I}_{0:t})}\\ &- \Delta t\frac{\sum_{\boldsymbol{y}'}\mathcal{R}(\boldsymbol{y}|\boldsymbol{y}')\mathbb{P}_t(\boldsymbol{y}'|\boldsymbol{I}_{0:t})}{\mathbb{P}_t(\boldsymbol{y}|\boldsymbol{I}_{0:t})}.\end{aligned}$$

Thus, by defining $L_t(\boldsymbol{x},\boldsymbol{y}) := \log\frac{\mathbb{P}_t(\boldsymbol{x}|\boldsymbol{I}_{0:t})}{\mathbb{P}_t(\boldsymbol{y}|\boldsymbol{I}_{0:t})}$, we have

$$\begin{aligned}\frac{\mathrm{d}L_t(\boldsymbol{x},\boldsymbol{y})}{\mathrm{dt}} \overset{\circ}{=}& \frac{1}{\mathbb{P}_t(\boldsymbol{x}|\boldsymbol{I}_{0:t})}\frac{\mathrm{d}\mathbb{P}_t(\boldsymbol{x}|\boldsymbol{I}_{0:t})}{\mathrm{dt}} - \frac{1}{\mathbb{P}_t(\boldsymbol{y}|\boldsymbol{I}_{0:t})}\frac{\mathrm{d}\mathbb{P}_t(\boldsymbol{y}|\boldsymbol{I}_{0:t})}{\mathrm{dt}}\\ \overset{\circ}{=}& \sum_{i\in\{1,2\}}[B_i + B_i^c] + B_\Upsilon\\ &+ \frac{\sum_{\boldsymbol{x}'}\mathcal{R}(\boldsymbol{x}|\boldsymbol{x}')\mathbb{P}_t(\boldsymbol{x}'|\boldsymbol{I}_{0:t})}{\mathbb{P}_t(\boldsymbol{x}|\boldsymbol{I}_{0:t})}\\ &- \frac{\sum_{\boldsymbol{y}'}\mathcal{R}(\boldsymbol{y}|\boldsymbol{y}')\mathbb{P}_t(\boldsymbol{y}'|\boldsymbol{I}_{0:t})}{\mathbb{P}_t(\boldsymbol{y}|\boldsymbol{I}_{0:t})},\end{aligned} \tag{2}$$

where $\overset{\circ}{=}$ represents the Stratonovich interpretation. Then, after the change of the variable, we can obtain

$$\begin{aligned}\frac{\mathrm{d}\mathbb{P}_t(\boldsymbol{x}|\boldsymbol{I}_{0:t})}{\mathrm{dt}} \overset{\circ}{=}& \mathbb{P}_t(\boldsymbol{x}|\boldsymbol{I}_{0:t})\sum_{\boldsymbol{x}'}\mathcal{B}(\boldsymbol{x},\boldsymbol{x}')\mathbb{P}_t(\boldsymbol{x}'|\boldsymbol{I}_{0:t})\\ &+ \sum_{\boldsymbol{x}'}\mathcal{R}(\boldsymbol{x}|\boldsymbol{x}')\mathbb{P}_t(\boldsymbol{x}'|\boldsymbol{I}_{0:t}),\end{aligned} \tag{3}$$

where, for $\boldsymbol{x},\boldsymbol{x}'\in\mathcal{X}_I$,

$$\mathcal{B}(\boldsymbol{x},\boldsymbol{x}') := \sum_{i\in\{1,2\}}[B_i(\boldsymbol{x},\boldsymbol{x}') + B_i^c(\boldsymbol{x},\boldsymbol{x}')] + B_\Upsilon(\boldsymbol{x},\boldsymbol{x}').$$

We can easily check that Eqs. (2) and (3) are equivalent as follows. From Eq. (3), we have

$$\begin{aligned}\frac{\frac{\mathrm{d}\mathbb{P}_t(\boldsymbol{x}|\boldsymbol{I}_{0:t})}{\mathrm{dt}}}{\mathbb{P}_t(\boldsymbol{x}|\boldsymbol{I}_{0:t})} \overset{\circ}{=}& \sum_{\boldsymbol{x}'}\mathcal{B}(\boldsymbol{x},\boldsymbol{x}')\mathbb{P}_t(\boldsymbol{x}'|\boldsymbol{I}_{0:t})\\ &+ \frac{\sum_{\boldsymbol{x}'}\mathcal{R}(\boldsymbol{x}|\boldsymbol{x}')\mathbb{P}_t(\boldsymbol{x}'|\boldsymbol{I}_{0:t})}{\mathbb{P}_t(\boldsymbol{x}|\boldsymbol{I}_{0:t})}.\end{aligned}$$

Obviously, the last two terms in Eq. (2) appear directly from the second term in the right-hand side of this equation as

$$\frac{\sum_{\boldsymbol{x}'}\mathcal{R}(\boldsymbol{x}|\boldsymbol{x}')\mathbb{P}_t(\boldsymbol{x}'|\boldsymbol{I}_{0:t})}{\mathbb{P}_t(\boldsymbol{x}|\boldsymbol{I}_{0:t})} - \frac{\sum_{\boldsymbol{y}'}\mathcal{R}(\boldsymbol{y}|\boldsymbol{y}')\mathbb{P}_t(\boldsymbol{y}'|\boldsymbol{I}_{0:t})}{\mathbb{P}_t(\boldsymbol{y}|\boldsymbol{I}_{0:t})}.$$

The first two terms in Eq. (2) can be calculated from Eq. (3) as

$$\begin{aligned}&\sum_{\boldsymbol{x}'}\mathcal{B}(\boldsymbol{x},\boldsymbol{x}')\mathbb{P}_t(\boldsymbol{x}'|\boldsymbol{I}_{0:t}) - \sum_{\boldsymbol{x}'}\mathcal{B}(\boldsymbol{y},\boldsymbol{x}')\mathbb{P}_t(\boldsymbol{x}'|\boldsymbol{I}_{0:t})\\ &\quad= \sum_{\boldsymbol{x}'}\left[\mathcal{B}(\boldsymbol{x},\boldsymbol{x}') - \mathcal{B}(\boldsymbol{y},\boldsymbol{x}')\right]\mathbb{P}_t(\boldsymbol{x}'|\boldsymbol{I}_{0:t})\\ &\quad= \sum_{\boldsymbol{x}'}\mathcal{B}(\boldsymbol{x},\boldsymbol{y})\mathbb{P}_t(\boldsymbol{x}'|\boldsymbol{I}_{0:t}) = \mathcal{B}(\boldsymbol{x},\boldsymbol{y}),\end{aligned}$$

where we use the equality that $\mathcal{B}(\boldsymbol{x},\boldsymbol{x}') - \mathcal{B}(\boldsymbol{y},\boldsymbol{x}') = \mathcal{B}(\boldsymbol{x},\boldsymbol{y})$.

Let $\boldsymbol{z}(t)\in[0,1]^{\#\mathcal{X}_I}$ be a vector representation of $\mathbb{P}_t(\boldsymbol{x}|\boldsymbol{I}_{0:t})$ for $\boldsymbol{x}\in\mathcal{X}_I$ as

$$\boldsymbol{z}_i(t) = \mathbb{P}_t(\chi_i\in\mathcal{X}_I|\boldsymbol{I}_{0:t}) \qquad \text{for } i\in\{1,\dots,\#\mathcal{X}_I\}.$$

Then, we have

$$\frac{\mathrm{d}\boldsymbol{z}(t)}{\mathrm{dt}} \overset{\circ}{=} \mathrm{diag}(\boldsymbol{z}(t))\mathcal{B}\boldsymbol{z}(t) + \mathcal{R}\boldsymbol{z}(t), \tag{4}$$

where $\mathrm{diag}(\boldsymbol{z}(t))$ is a diagonal matrix generated by a vector $\boldsymbol{z}(t)$. $\mathcal{B}$ and $\mathcal{R}$ are matrix representations of $\mathcal{B}(\boldsymbol{x},\boldsymbol{y})$ and $A(\boldsymbol{x},\boldsymbol{y})$ for $\boldsymbol{x},\boldsymbol{y}\in\mathcal{X}_I$. It should be noted that Eq. (4) is interpreted as

$$\mathrm{d}\boldsymbol{z}(t) \overset{\circ}{=} \left[\mathrm{diag}(\boldsymbol{z}(t))\mathcal{B}\boldsymbol{z}(t) + \mathcal{R}\boldsymbol{z}(t)\right]\mathrm{dt}.$$

B. Bayesian gates in Ito form

In order to numerically calculate Eq. (4), its Ito form is convenient. To obtain the Ito form of Eq. (4), let us substitute $I_i(t)\mathrm{d}t = \nu_i(\boldsymbol{x}(t))\mathrm{d}t + \sigma_i \mathrm{d}W_t^i$ as

$$\begin{aligned} B_i(\boldsymbol{x},\boldsymbol{y})\mathrm{d}t &= \frac{\Delta\nu_i(I_i(t) - \langle\nu\rangle_i)}{\bar{\sigma}_i^2}\mathrm{d}t \\ &= \frac{\Delta\nu_i[(\nu_i(t) - \langle\nu\rangle_i)\mathrm{d}t + \sigma_i \mathrm{d}W_t^i]}{\bar{\sigma}_i^2}, \end{aligned}$$

and

$$\begin{aligned} B_i^c(\boldsymbol{x},\boldsymbol{y})\mathrm{d}t &= -\frac{\rho\Delta\nu_i I_i(t)}{\bar{\sigma}_1\bar{\sigma}_2}\mathrm{d}t \\ &= -\frac{\rho\Delta\nu_i}{\bar{\sigma}_1\bar{\sigma}_2}(\nu_i(t)\mathrm{d}t + \sigma_i \mathrm{d}W_t^i), \end{aligned}$$

where $\nu_i(t) = \nu_i(\boldsymbol{x}(t))$. Then, let us define

$$\begin{aligned} B_i^I(\boldsymbol{x},\boldsymbol{y}) &:= \Delta\nu_i\left[\frac{(\nu_i(t) - \langle\nu\rangle_i)}{\bar{\sigma}_i^2} - \frac{\rho\bar{\nu}_i(t)}{\bar{\sigma}_1\bar{\sigma}_2}\right] \\ B_i^W(\boldsymbol{x},\boldsymbol{y}) &:= \Delta\nu_i\left[\frac{1}{\bar{\sigma}_i} - \frac{\rho\sigma_i}{\bar{\sigma}_1\bar{\sigma}_2}\right], \end{aligned}$$

and then $\mathcal{B}$ becomes

$$\mathcal{B}(\boldsymbol{x},\boldsymbol{y})\mathrm{d}t = \left[B^I + B_\Upsilon\right]\mathrm{d}t + \sum_i B_i^W \mathrm{d}W_t^i,$$

where $B^I := \sum_{i\in\{1,2\}} B_i^I$, and we have

$$\begin{aligned} \mathrm{d}\boldsymbol{z}_t = &\left(\mathrm{diag}(\boldsymbol{z}_t)\left[B^I + B_\Upsilon\right]\boldsymbol{z}_t + \mathcal{R}\boldsymbol{z}_t\right)\mathrm{d}t \\ &+ \sum_i \mathrm{diag}(\boldsymbol{z}_t)B_i^W\boldsymbol{z}_t \circ \mathrm{d}W_t^i, \end{aligned}$$

where $\boldsymbol{z}_t = \boldsymbol{z}(t)$. Thus, the Ito version of this Stratonovich stochastic differential equation becomes

$$\begin{aligned} \mathrm{d}\boldsymbol{z}_t = &\Big[\left(\mathrm{diag}(\boldsymbol{z}_t)\left[B^I + B_\Upsilon\right]\boldsymbol{z}_t + \mathcal{R}\boldsymbol{z}_t\right) \\ &+ \frac{1}{2}\sum_i (\nabla_z \mathrm{diag}(\boldsymbol{z}_t)B_i^W\boldsymbol{z}_t)(\mathrm{diag}(\boldsymbol{z}_t)B_i^W\boldsymbol{z}_t))\Big]\mathrm{d}t \\ &+ \sum_i \mathrm{diag}(\boldsymbol{z}_t)B_i^W\boldsymbol{z}_t\mathrm{d}W_t^i. \end{aligned}$$

By expanding the matrices, we have

$$\begin{aligned} \left[\mathrm{diag}(\boldsymbol{z}_t)B_i^W\boldsymbol{z}_t\right]_k =& (z_t)_k \sum_j B_{k,j;i}^W (z_t)_j \\ \nabla_z \mathrm{diag}(\boldsymbol{z}_t)B_i^W\boldsymbol{z}_t =& \begin{pmatrix} (\boldsymbol{z}_t)_1 B_{1,1;i}^W & \cdots & \cdots \\ \vdots & (\boldsymbol{z}_t)_2 B_{2,2;i}^W & \cdots \\ \vdots & \vdots & \cdots \end{pmatrix} \\ &+ \begin{pmatrix} (\boldsymbol{z}_t)_1 B_{1,1;i}^W & (\boldsymbol{z}_t)_1 B_{1,2;i}^W & \cdots \\ (\boldsymbol{z}_t)_2 B_{2,1;i}^W & (\boldsymbol{z}_t)_2 B_{2,2;i}^W & \cdots \\ \vdots & \vdots & \cdots \end{pmatrix} \\ =& \mathrm{diag}(B_i^W\boldsymbol{z}_t) + \mathrm{diag}(\boldsymbol{z}_t)B_i^W \end{aligned}$$

Finally, we have the Ito form as

$$\begin{aligned} \frac{\mathrm{d}\boldsymbol{z}(t)}{\mathrm{d}t} =& \Big[(\mathrm{diag}(\boldsymbol{z}(t))\mathcal{B}\boldsymbol{z}(t) + \mathcal{R}\boldsymbol{z}(t)) \\ &+ \frac{1}{2}\sum_i \left[\mathrm{diag}(B_i^W\boldsymbol{z}(t)) + \mathrm{diag}(\boldsymbol{z}(t))B_i^W\right] \\ &\times (\mathrm{diag}(\boldsymbol{z}(t))B_i^W\boldsymbol{z}(t))\Big]\mathrm{d}t \\ &+ \sum_i \mathrm{diag}(\boldsymbol{z}_t)B_i^W\boldsymbol{z}_t\mathrm{d}W_t^i. \end{aligned}$$

C. Derivation of Bayesian NOR, XOR, and AND gates

Let consider the situation where there is one gate input $I_1(t)$, as in Fig. 1(A) in the main text, which is realized by setting $\nu_2 = 0$, $\sigma_2 = 0$, and $\rho = 0$. We set $\sigma_1 = \sigma_0$. In addition, we assume that ν_1 in I_1 is as

$$\nu_1(\boldsymbol{x}(t)) = U_1(x_1(t)) + U_2(x_2(t)),$$

where

$$U_1(x) = U_2(x) = \begin{cases} \mu/2 & \text{for } x = 1 \\ -\mu/2 & \text{for } x = 0 \end{cases}.$$

Because of this definition of $\nu_1(\boldsymbol{x}(t))$, $\mathcal{X}_I$ has three subsets of $\mathcal{X}$, $\chi_1 = \{(0,0)\}$, $\chi_2 = \{(1,0),(0,1)\}$, and $\chi_3 = \{(1,1)\}$. For each χ_i, we assign z_i. Under these conditions, we have

$$B_1(\boldsymbol{x},\boldsymbol{y}) := \frac{(\nu_1(\boldsymbol{x}) - \nu_1(\boldsymbol{y}))(I_1(t') - (\nu_1(\boldsymbol{x}) + \nu_1(\boldsymbol{y}))/2)}{\sigma_0^2},$$

$B_2(\boldsymbol{x},\boldsymbol{y}) = 0$, $B_i^c(\boldsymbol{x},\boldsymbol{y}) = 0$, and $B_\Upsilon(\boldsymbol{x},\boldsymbol{y}) = 0$ for $\boldsymbol{x},\boldsymbol{y} \in \mathcal{X}_I$. Since we have $\nu_1(\chi_1) = -\mu$, $\nu_1(\chi_2) = 0$, and $\nu_1(\chi_3) = \mu$,

$$\begin{aligned} \mathcal{B} &= \frac{\mu}{\sigma_0^2}\begin{pmatrix} 0 & -(I_1(t') + \frac{\mu}{2}) & -2I_1(t') \\ (I_1(t') + \frac{\mu}{2}) & 0 & -(I_1(t') - \frac{\mu}{2}) \\ 2I_1(t') & (I_1(t') - \frac{\mu}{2}) & 0 \end{pmatrix}, \\ &= \frac{\mu}{\sigma_0^2}\left[I_1(t')\begin{pmatrix} 0 & -1 & -2 \\ 1 & 0 & -1 \\ 2 & 1 & 0 \end{pmatrix} + \frac{\mu}{2}\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}\right]. \end{aligned}$$

The transition matrix $\mathcal{R}$ can be obtained by considering the transition rate of individual $x_1(t)$ and $x_2(t)$. When the transition rates from 0 to 1 and from 1 to 0 are r_{on}^i and r_{off}^i for each $i \in \{1,2\}$, we have the following transition rates between $\boldsymbol{x},\boldsymbol{x}' \in \mathcal{X}$:

$$\tilde{\mathcal{R}} = \begin{pmatrix} -r_{\mathrm{on}}^1 - r_{\mathrm{on}}^2 & r_{\mathrm{off}}^2 & r_{\mathrm{off}}^1 & 0 \\ r_{\mathrm{on}}^1 & -r_{\mathrm{on}}^1 - r_{\mathrm{off}}^2 & 0 & r_{\mathrm{off}}^1 \\ r_{\mathrm{on}}^2 & 0 & -r_{\mathrm{on}}^2 - r_{\mathrm{off}}^1 & r_{\mathrm{off}}^2 \\ 0 & r_{\mathrm{on}}^1 & r_{\mathrm{on}}^2 & -r_{\mathrm{off}}^1 - r_{\mathrm{off}}^2 \end{pmatrix},$$

where we should note that $\mathcal{X} = \{(0,0),(0,1),(1,0),(1,1)\}$. By setting $r_{\mathrm{on}}^i = r_{\mathrm{on}}$

and $r^i_{\rm off} = r_{\rm off}$, $\tilde{\mathcal{R}}$ is reduced to

$$\tilde{\mathcal{R}} = \begin{pmatrix} -2r_{\rm on} & r_{\rm off} & r_{\rm off} & 0 \\ r_{\rm on} & -r_{\rm on} - r_{\rm off} & 0 & r_{\rm off} \\ r_{\rm on} & 0 & -r_{\rm on} - r_{\rm off} & r_{\rm off} \\ 0 & r_{\rm on} & r_{\rm on} & -2r_{\rm off} \end{pmatrix}.$$

By degenerating $\mathcal{X}$ into $\mathcal{X}_I = \{\{(0,0)\},\{(0,1),(0,1)\},\{(1,1)\}\}$ where $\boldsymbol{x} = (0,1)$ and $\boldsymbol{x}' = (0,1)$ are identified, we obtain $\mathcal{R}$ in the main text:

$$\mathcal{R} = \begin{pmatrix} -2r_{\rm on} & r_{\rm off} & 0 \\ 2r_{\rm on} & -r_{\rm on} - r_{\rm off} & 2r_{\rm off} \\ 0 & r_{\rm on} & -2r_{\rm off} \end{pmatrix}.$$

[1] A. Eldar and M. B. Elowitz, Nature **467**, 167 (2010).
[2] A. A. Faisal, L. P. J. Selen, and D. M. Wolpert, Nature Reviews Neuroscience **9**, 292 (2008).
[3] O. H. Schmitt, Journal of Scientific Instruments **15**, 24 (2002).
[4] K. Murali, S. Sinha, W. L. Ditto, and A. R. Bulsara, Phys. Rev. Lett. **102**, 104101 (2009).
[5] K. Murali, I. Rajamohamed, S. Sinha, W. L. Ditto, and A. R. Bulsara, Applied Physics Letters **95** (2009).
[6] S. Sinha, J. M. Cruz, T. Buhse, and P. Parmananda, Europhys. Lett. **86**, 60003 (2009).
[7] D. C. Knill and A. Pouget, Trends Neurosci. **27**, 712 (2004).
[8] T. J. Kobayashi, Phys. Rev. Lett. **104**, 228104 (2010).
[9] T. J. Kobayashi, Phys. Rev. Lett. **106**, 228101 (2011).
[10] B. Nazer and M. Gastpar, IEEE Trans. Inform. Theory **53** (2007).
[11] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis* (Springer, 1985).
[12] C. W. Gardiner, *Handbook of stochastic methods for physics, chemistry, and the natural sciences* (1985).
[13] L. Gammaitoni, Applied Physics Letters **91**, 224104 (2007).
[14] S. Navlakha and Z. Bar-Joseph, Mol Syst Biol **7** (2011).
[15] P. E. M. Purnick and R. Weiss, Nat Rev Mol Cell Biol **10**, 410 (2009).