Nano-Optics and Nanophotonics

Makoto Naruse *Editor*

Nanophotonic Information Physics

Nanointelligence and Nanophotonic Computing

Springer

Nano-Optics and Nanophotonics

For further volumes:
http://www.springer.com/series/8765

The Springer Series in Nano-Optics and Nanophotonics provides an expanding selection of research monographs in the area of nano-optics and nanophotonics, science- and technology-based on optical interactions of matter in the nanoscale and related topics of contemporary interest. With this broad coverage of topics, the series is of use to all research scientists, engineers and graduate students who need up-to-date reference books. The editors encourage prospective authors to correspond with them in advance of submitting a manuscript. Submission of manuscripts should be made to the editor-in-chief, one of the editors or to Springer.

Makoto Naruse
Editor

Nanophotonic Information Physics

Nanointelligence and Nanophotonic Computing

Springer

Editor
Makoto Naruse
Photonic Network Research Institute
National Institute of Information and
Communications Technology
Tokyo
Japan

ISSN 2192-1970 ISSN 2192-1989 (electronic)
ISBN 978-3-642-40223-4 ISBN 978-3-642-40224-1 (eBook)
DOI 10.1007/978-3-642-40224-1
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013953214

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This book contains nine review articles about nanophotonic information physics—with the subtitle nanointelligence and nanophotonic computing—and is the first publication of its kind.

Optical science and technologies have been experiencing tremendous advancements in recent years, especially at scales below the wavelength of light. Technological enablers have been cultivated both in top-down approaches such as nanoscale-precision lithography, and bottom-up ones, such as DNA-based self-assembly. The fundamental physical processes of light–matter interactions, such as energy transfer via near-field interactions, among others, have been extensively studied from the viewpoint of the basic science and enabling technologies of nanophotonics.

From an information standpoint, on the other hand, novel architectures should be considered so that we can fully utilize and benefit from the potential of the unique physical processes made possible by nanophotonics. There are plenty of degrees-of-freedom on the nanoscale, and a wide variety of physical processes and technologies exist. A comprehensive view covering diverse disciplines, including physics, materials, devices, systems, information, and architectures, is vitally important and will push the frontiers of physical and information sciences and technologies.

At the same time, such interdisciplinary and basic research areas that cover both physics and information—or real-world and abstract functions—will pave the way to gain fundamental insights and critical knowledge for implementing novel applications of nanophotonics that will be of key importance in our lives and for wider society, both today and in the future; these include computing, information systems, network systems, sensing and imaging, healthcare and welfare, safety and security, the environment and energy.

Such a research concept is represented by the title of this book. The nine selected articles in this book are based on presentations given at the First International Workshop on Information Physics and Computing in Nano-scale Photonics (IPCN), held in Orleans, France, in September 2012. The workshop aimed to bring researchers together and stimulate strong interest at the intersection of nanophotonics, nanoelectronics, and information science and technologies, a goal that is also shared by this book. It should be emphasized that novel architectural

ideas, concepts, and paradigms are discussed in each chapter, besides concrete technological realizations.

Chapter 1 (Naruse et al.) discusses information physics fundamentals for nanophotonics. Chapters 2 (Dwyer et al.), 3 (Tanida), and 4 (Ogura et al.) present the original concepts of the authors and some practical applications. The concepts described in these three chapters are all based on DNA and related self-assembly principles and technologies, but it should be remarked that a variety of system-level ideas and benefits result. Chapters 5 (Kasai et al.) and 6 (Kawahito et al.) are based on nanoelectronics with unique device architectures for computing and imaging applications. Chapter 7 (Tate et al.) discusses interfacing issues, and corresponding solutions, between the macro- and nano-scale worlds, which are of critical importance for nanophotonic systems. Chapter 8 (Tait et al.) discusses photonic neuromorphic signal processing and computing, including both its fundamental theoretical concerns and photonic realizations. Chapter 9 (Aono et al.) paves the way to new problem-solving and decision-making methods based on the spatiotemporal optical excitation transfer dynamics provided by optical near-field interactions.

Through these inter- and cross-disciplinary investigations covering optical, material, and information sciences, besides considering applications and architectures that will provide new value via state-of-the-art technologies, it is my sincere hope to convey the excitement and sense-of-wonder that emerges, and that researchers will harness the fruits of these investigations in creating a new research area that merges physical and information sciences and technologies.

I am deeply indebted to the committee members, speakers, and participants of the IPCN Workshop and its subsequent in-depth discussions, especially Dr. Y. Ogura (Osaka University) who served as the co-chair of the workshop, and Dr. M. Aono (Tokyo Institute of Technology) who greatly helped in the successful organization of the workshop. Also, I would like to thank Dr. J. Tanida (Osaka University) for his encouragement and discussions throughout the course of the workshop and the preparation of this book. In addition, I sincerely thank Dr. M. Ohtsu (The University of Tokyo), who is the Editor-in-Chief of Nano-Optics and Nanophotonics, for his great encouragement of researchers in nanophotonics science and technology and for his support and suggestions for this book. Finally, I am grateful to Dr. C. Acheron of Springer–Verlag for his guidance and suggestions throughout the preparation of this book.

Tokyo Makoto Naruse

Contents

Contributors

Masashi Aono Earth-Life Science Institute, Tokyo Institute of Technology, 201201 Ookayama, Meguro-ku, Tokyo 152-8850, Japan, e-mail: masashi.aono@elsi.jp

Chris Dwyer Department of Electrical and Computer Engineering, Department of Computer Science, Duke University, Durham, NC, USA, e-mail: dwyer@ece.duke.edu

Masahiko Hara Department of Electronic Chemistry, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama 226-8503, Japan, e-mail: masahara@echem.titech.ac.jp

Hirokazu Hori Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, 4-3-11 Takeda, Kofu, Yamanashi 400-8511, Japan, e-mail: hirohori@yamanashi.ac.jp

Keiichiro Kagawa Research Institute of Electronics, Shizuoka University, Hamamatsu, Japan, e-mail: kagawa@idl.rie.shizuoka.ac.jp

Seiya Kasai Graduate School of Information Science and Technology and Research Center for Integrated Quantum Electronics, Hokkaido University, N14, W9, Sapporo 060-0814, Japan, e-mail: kasai@rciqe.hokudai.ac.jp

Shoji Kawahito Research Institute of Electronics, Shizuoka University, Hamamatsu, Japan, e-mail: kawahito@idl.rie.shizuoka.ac.jp

Song-Ju Kim Atomic Electronics Group, WPI Center for Materials Nanoarchitectonics (MANA), National Institute for Materials Science (NIMS), 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan, e-mail: KIM.Songju@nims.go.jp

Tarmer Mohamed Department of Electrical and Computer Engineering, University of Calgary, Calgary, Alberta, T2N 1N4, Canada, e-mail: tamer.mohamed@ucalgary.ca

Mohammad Mottaghi Department of Computer Science, Duke University, Durham, NC, USA, e-mail: mamad@cs.duke.edu

Mitchell A. Nahmias Lightwave Communications Laboratory, Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA, e-mail: mnahmias@princeton.edu

Makoto Naruse Photonic Network Research Institute, National Institute of Information and Communications Technology, 4-2-1 Nukui-kita, Koganei, Tokyo 184-8795, Japan, e-mail: naruse@nict.go.jp

Takahiro Nishimura Graduate School of Information Science and Technology, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan, e-mail: t-nishimura@ist.osaka-u.ac.jp

Yusuke Ogura Graduate School of Information Science and Technology, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan, e-mail: ogura@ist.osaka-u.ac.jp

Motoichi Ohtsu Department of Electrical Engineering and Information Systems and Nanophotonics Research Center, Graduate School of Engineering, The University of Tokyo, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-8656, Japan, e-mail: ohtsu@ee.t.u-tokyo.ac.jp

Paul R. Prucnal Lightwave Communications Laboratory, Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA, e-mail: prucnal@princeton.edu

Arjun Rallapalli Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA, e-mail: ar233@duke.edu

Bhavin J. Shastri Lightwave Communications Laboratory, Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA, e-mail: bshastri@princeton.edu

Yuta Shiratori Graduate School of Information Science and Technology and Research Center for Integrated Quantum Electronics, Hokkaido University, N14, W9, Sapporo 060-0814, Japan

Alexander N. Tait Lightwave Communications Laboratory, Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA, e-mail: atait@princeton.edu

Jun Tanida Graduate School of Information Science and Technology, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan, e-mail: tanida@ist.osaka-u.ac.jp

Naoya Tate Department of Electrical Engineering and Information Systems and Nanophotonics Research Center, Graduate School of Engineering, The University of Tokyo, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-8656, Japan, e-mail: tate@nanophotonics.t.u-tokyo.ac.jp

Yue Tian Lightwave Communications Laboratory, Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA, e-mail: yuetian@princeton.edu

Masamitsu Wakabayashi Department of Biomolecular Engineering, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama 226-8501, Japan, e-mail: wakabayashi@riken.jp

Siyang Wang Department of Electrical and Computer Engineering, Department of Computer Science, Duke University, Durham, NC, USA, e-mail: siyang.wang@duke.edu

Kenji Yamada Graduate School of Medicine, Osaka University, 1-7 Yamadaoka, Suita, Osaka 565-0871, Japan, e-mail: k-yamada@sahs.med.osaka-u.ac.jp

Hirotsugu Yamamoto Department of Optical Science and Technology, The University of Tokushima, 2-1 Minamijyousanjima-cho, Tokushima 770-8506, Japan, e-mail: yamamoto@opt.tokushima-u.ac.jp

Svetlana N. Yanushkevich Department of Electrical and Computer Engineering, University of Calgary, Calgary, Alberta, T2N 1N4, Canada, e-mail: syanshk@ucalgary.ca

Keita Yasutomi Research Institute of Electronics, Shizuoka University, Hamamatsu, Japan, e-mail: kyasu@idl.rie.shizuoka.ac.jp

Hong-Quan Zhao Graduate School of Information Science and Technology and Research Center for Integrated Quantum Electronics, Hokkaido University, N14, W9, Sapporo 060-0814, Japan

Chapter 1
Nanointelligence: Information Physics Fundamentals for Nanophotonics

Makoto Naruse, Naoya Tate, Masashi Aono and Motoichi Ohtsu

Abstract Nanophotonics has been extensively studied with the aim of unveiling and exploiting light-matter interactions that occur at a scale below the diffraction limit of light. From the viewpoint of information, novel architectures, novel design and analysis principles, and even novel computing paradigms should be considered so that we can fully benefit from the potential of nanophotonics for various applications. In this chapter, we first present some fundamental and emergent attributes associated with optical excitation transfer mediated by optical near-field interactions. Toward achieving a computing paradigm that surpasses the classical von Neumann architecture, we describe stochastic solution searching, which exploits the spatiotemporal dynamics of optical excitation transfer. Second, we show information security applications based on near-field applications, together with their theoretical and experimental foundations. Finally, we present a stochastic analysis of light-assisted self-organized material formation in order to gain a deeper understanding of the underlying physics. We consider that a common feature across all of these demonstrations is the extraction of "intelligent" functions and behaviors from an information-based standpoint—an idea represented by the title of this chapter, "*nanointelligence*".

M. Naruse (✉)
Photonic Network Research Institute, National Institute of Information and Communications Technology, 4-2-1 Nukui-kita, Koganei, Tokyo 184-8795, Japan
e-mail: naruse@nict.go.jp

N. Tate · M. Ohtsu
Department of Electrical Engineering and Information Systems and Nanophotonics Research Center, Graduate School of Engineering, The University of Tokyo, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-8656, Japan

M. Aono
Earth-Life Science Institute, Tokyo Institute of Technology, 201202 Ookayama, Meguro-ku, Tokyo 152-8550, Japan
e-mail: masashi.aono@elsi.jp

M. Naruse (ed.), *Nanophotonic Information Physics*,
Nano-Optics and Nanophotonics, DOI: 10.1007/978-3-642-40224-1_1,

1.1 Introduction

Light plays a crucial role in information devices and systems in a range of fields, for example, communication, information processing, imaging, and displays. There is no doubt that the superior physical attributes of light constitute the foundations of the benefits for which we are greatly indebted today. At the same time, however, ever increasing quantitative demands, such as the massive amount of digital information carried on networks, necessitate further advancements in optics and photonics. In addition, there has also been a surge in demand for qualitatively novel technologies, such as energy saving [1], mobile and ubiquitous devices [2], solid-state lighting and displays, healthcare and welfare, and safety and security [3, 4], to name a few. Conventional optics and photonics, however, suffer from difficulties in resolving these quantitative and qualitative challenges due to their fundamental limitations, principally, the diffraction limit of light [5].

Nanophotonics, which makes use of interactions between light and matter at a scale below the wavelength of light, has advanced remarkably [6, 7]. An optical near-field, which is localized in the vicinity of nanostructured matter, is free of the diffraction effects imposed on propagating light, in a sense breaking through the diffraction limit, which allows the physical extent of light itself to be reduced [8, 9]. Furthermore, the nature of photons on the nanometer scale, including higher-order atom–light interactions [10], has led to the discovery of unique phenomena observed only on the nanoscale, revealing a physical picture of "dressed photons", or photons dressed by material excitations [11, 12]. Thus, not only does nanophotonics possess the ability to break through the diffraction limit, but it also allows physical processes that are unachievable conventionally, such as dipole-forbidden transitions. These novel optical near-field processes have led to various devices that have been experimentally demonstrated recently, including light concentration [13], infrared-to-visible light conversion [14], silicon light emission [15, 16], solar cells [17], and so forth. The rapid progress of experimental technologies has been a driving force behind the advances in nanophotonics; ultrafast spectroscopy for nanostructures [18], and size- and position-controlled quantum nanostructures, such as InAs [19], ZnO [20], and shape-engineered nanostructures [21–23], among others.

From the viewpoint of information or system design, on the other hand, there are many unresolved, important basic issues in nanophotonics. For example, system architectures, basic structures for achieving versatile functions, modeling, design, and analysis principles and methods that inherit the physical principles of nanophotonics should be developed. This chapter sheds light on some system-level fundamentals and insights in nanophotonics, with particular focus on achieving "*intelligent*" functions, what we summarize as the concept called "*nanointelligence*".

It should be emphasized that replacing or competing with conventional computing and information technologies is not necessarily the primary motivation. What should be pursued is to exploit and maximize the potential of the unique physical attributes inherent in nanophotonics. Nanophotonic security, discussed in Sect. 1.3, is one example application where accessibility via light is essential, and optical

near-field processes provide unique solutions [3]. Furthermore, insights gained in the modeling and analysis of nanophotonics are applicable to a wide range of applications besides computing or information processing. They can also contribute to design and analysis in energy applications [24] and nanofabrication [25]. Dwyer et al. have been developing a nano-optical computer architecture for medical and healthcare applications based on resonant energy transfer made possible by DNA-based self-assembly [26, 27]. Catrysse et al. have investigated nanostructures for imaging applications [28], where light–matter interactions are essentially and inevitably present, and the unique nanophotonic solutions offered are valuable. Furthermore, system-oriented approaches are beneficial in investigating material formation. Stochastic modeling and analysis in optical-near-field-assisted nanofabrication, which will be dealt with in Sect. 1.4, are typical examples. In more general contexts, dealing with light and matter in a nonequilibrium open system [29] would provide new knowledge in nanophotonics too, as in other fields. Self-organized criticality [30], observed in some nanophotonic experiments [31, 32], is a natural emergent property thanks to the inclusion of energy flow in addition to elemental near-field interactions.

Summing up all of these related aspects, that is to say, a system-oriented approach, or an approach toward realizing intelligent functions, this chapter reviews nanophotonics from an information physics approach, in particular, by examining the three aspects described below.

Optical Excitation Transfer to Go Beyond the von Neumann Architecture
In Sect. 1.2, we deal with localized optical excitation transfer. Conventionally, propagating light is assumed to interact with nanostructured matter in a spatially uniform manner—a principle referred to as the long-wavelength approximation—from which state transition rules are derived, including dipole-forbidden transitions. However, such an approximation is not valid in the case of a localized optical near-field in the vicinity of nanostructured matter; the inhomogeneity of localized light makes even conventionally dipole-forbidden transitions allowable [5]. We will first review some basic and unique attributes of optical excitation transfer from space-, time-, and energy-related perspectives, followed by a stochastic solution search, which will pave the way to a new computing paradigm beyond von Neumann architecture.

Nanophotonics for Security
Section 1.3 deals with security applications of nanophotonics. Optics has been applied in a variety of security applications [4]. However, since the fundamental physical principle usually involves optical far-fields, diffraction of light causes severe difficulties in device scaling and system integration, for example. Nanophotonics can break through the diffraction limit of conventional light and can provide additional functions, such as adding hidden information that is only retrievable via optical near-fields. Section 1.3 first characterizes one fundamental security feature via a rigorous nano-optical theory based on an angular-spectrum representation of electromagnetic fields on the nanoscale. Its applications to information hiding and authentication functions based on shape-engineering of nanostructures are demonstrated. The study will contribute to tamper-resistant hardware based on unique nano-optical physical processes available on the nanoscale.

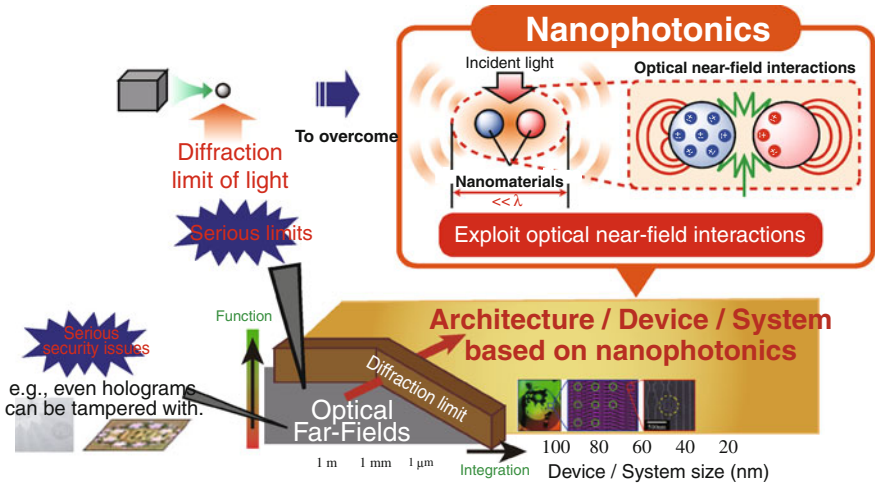


Fig. 1.1 Overall concept of the discussion presented in this chapter

Stochastic Modeling of Near-Field Processes for Intelligent Material Formation Nanofabrication involving optical near-field processes has demonstrated nanoparticle size regulation [33], nanoparticle array formation [34], the appearance of interesting photosensitivity [17], atomic-scale surface flattening [35], etc., where self-organizing, or "*intelligent*", behavior emerges by introducing light in material formation. Section 1.4 concerns stochastic modeling of material formation involving optical near-field processes that reproduce phenomenological characteristics consistent with the experimental observations, in order to gain a deeper understanding of the underlying physical mechanisms and to enable optimization of future devices.

The above concept is schematically shown in Fig. 1.1. Section 1.5 summarizes this chapter and discusses some future prospects.

1.2 Optical Excitation Transfer to go Beyond the von Neumann Architecture

In this section, we discuss optical excitation transfer involving optical near-field interactions. In the literature, dipole–dipole interactions, such as Förster resonant energy transfer, are typically referred to in explaining energy transfer from smaller quantum dots (QDs) to larger ones [36, 37]. However, it should be noted that such modeling based on point dipoles does not allow optical transitions to dipole-forbidden energy sublevels. Also, recent experimental observations in light harvesting antenna indicate the inaccuracy of dipole-based modeling [38–40]. On the other hand, as discussed below, the localized nature of optical near-fields frees us from conventional optical selection rules, meaning that optical excitations could excite QDs to energy levels

that are conventionally electric-dipole forbidden. Section 1.2.1 reviews the theoretical formalisms of optical excitation transfer. Section 1.2.2 examines some basic and unique attributes provided by optical excitation transfer from space-, time-, and energy-related perspectives. Section 1.2.3 discusses stochastic solution searching, paving the way to a new computing paradigm beyond the classical von Neumann architecture.

1.2.1 Fundamentals

We begin with the interaction Hamiltonian between an electron–hole pair and an electric field, which is given by

$$\hat{H}_{int} = -\int d^3r \sum_{i,\ j=e,\ h} \hat{\psi}_i^{\dagger}(\mathbf{r}) e\mathbf{r} \bullet \mathbf{E}(\mathbf{r}) \hat{\psi}_j(\mathbf{r}), \tag{1.1}$$

where e represents the electron charge, $\hat{\psi}_i^{\dagger}(\mathbf{r})$ and $\hat{\psi}_j(\mathbf{r})$ are respectively creation and annihilation operators of either an electron ($i,\ j = e$) or a hole ($i,\ j = h$) at position $\mathbf{r}$, and $\mathbf{E}(\mathbf{r})$ is the electric field [24]. In usual light–matter interactions, $\mathbf{E}(\mathbf{r})$ is a constant since the electric field of propagating light is homogeneous on the nanometer scale. Therefore, we can derive optical selection rules by calculating the dipole transition matrix elements. As a consequence, in the case of spherical quantum dots, for instance, only transitions to states specified by $l = m = 0$ are allowed, where l and m are the orbital angular momentum quantum number and magnetic quantum number, respectively. In the case of optical near-field interactions, on the other hand, due to the large spatial inhomogeneity of the localized optical near-fields at the surface of nano-scale material, an optical transition that violates conventional optical selection rules is allowed. Detailed theory can be found in [12].

Here we assume two spherical quantum dots whose radii are R_{S} and R_{L}, which we call QD_S and QD_L, respectively, as shown in Fig. 1.2a. The energy eigenvalues of states specified by quantum numbers (n, l) are given by

$$E_{nl} = E_g + E_{ex} + \frac{\hbar^2 \alpha_{nl}^2}{2MR^2}\ (n = 1, 2, 3, \cdots), \tag{1.2}$$

where E_g is the band gap energy of the bulk semiconductor, E_{ex} is the exciton binding energy in the bulk system, M is the effective mass of the exciton, and α_{nl} are determined from the boundary conditions, for example, as $\alpha_{n0} = n\pi$, $\alpha_{11} = 4.49$. According to (1.2), there exists a resonance between the level of quantum number $(1, 0)$ in QD_S and that of quantum number $(1, 1)$ in QD_L if $R_{\mathrm{L}}/R_{\mathrm{S}} = 4.49/\pi \approx 1.43$. Note that the (1, 1)-level in QD_L is a dipole-forbidden energy level. However, an optical near-field, denoted by U_{SL_2} in Fig. 1.2a, allows this level to be populated due to the steep electric field in the vicinity of QD_S. Therefore, an exciton in the (1, 0)-level in QD_S could be transferred to the (1, 1)-level in QD_L. In QD_L, the excitation

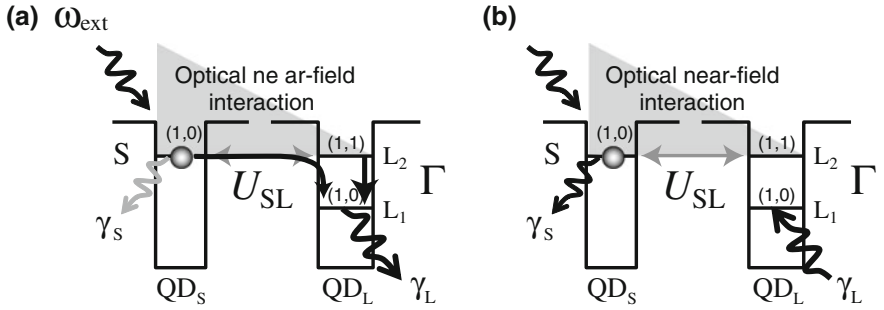


Fig. 1.2 **a** Optical excitation transfer from a smaller quantum dot (QD_S) to a larger one (QD_L) mediated by optical near-field interactions. Note that transition to the energy level L_2 in QD_L is conventionally dipole-forbidden.The energy dissipation occurring at QD_L, or the relaxation from L_2 to L_1, guarantees unidirectionality. **b** State filling induced at the lower energy level in the larger dot results in different flows of optical excitation

undergoes intersublevel energy relaxation due to exciton–phonon coupling, denoted by Γ, which is faster than the near-field interaction [41, 42], and the excitation relaxes to the (1, 0)-level, from where it radiatively decays. Also, since the radiation lifetime of quantum dots is inversely proportional to their volume [43], finally we find unidirectional optical excitation transfer from QD_S to QD_L.

We first introduce quantum mechanical modeling of the total system based on a density matrix formalism. There are in total eight states where either zero, one, or two excitation(s) can sit in the energy levels S, L_1, and L_2 in the system. The interactions between QD_S and QD_L are denoted by U_{SL}, and the radiative relaxation rates from S and L_1 are respectively given by γ_S and γ_L. The model Hamiltonian of the coupled two-dot system is given by

$$H = \hbar \begin{pmatrix} \Omega_S & U_{SL} \\ U_{SL} & \Omega_L \end{pmatrix} \tag{1.3}$$

where $\hbar U_{SL}$ is the optical near-field interaction, and $\hbar\Omega_S$ and $\hbar\Omega_L$ respectively refer to the eigenenergies of QD_S and QD_L. The behavior of optical excitation transfer is obtained by solving the equation of motion, given by the Liouville equation [44]

$$\dot{\rho}(t) = -\frac{i}{\hbar}[H, \rho(t)] - N_\Gamma \rho(t) - \rho(t) N_\Gamma, \tag{1.4}$$

where ρ is the density operator, and N_Γ is a diagonal matrix whose diagonal elements are $\gamma_S/2$ and $\Gamma/2$.

If necessary, we can explicitly include an external Hamiltonian H_{ext} representing the interaction between the external input light at frequency ω_{ext} and the quantum dot system; this is given by

$$H_{ext}(t) = gate(t) \times \Big[\Big(\exp(i(\Omega_S - \omega_{ext})S^{\dagger} + \exp(-i(\Omega_S - \omega_{ext})S\Big) + \Big(\exp(i(\Omega_{L_1} - \omega_{ext}))L_1^{\dagger} + \exp(-i(\Omega_{L_1} - \omega_{ext}))L_1\Big)\Big], \quad (1.5)$$

where $gate(t)$ specifies the duration and the amplitude of the external input light, and $S^{\dagger}$(S) and $L^{\dagger}(L)$ represents creation (annihilation) operators regarding the (1, 0)-levels in QD_S and QD_L, respectively. If control light occupies the energy level L_1, namely, state filling is induced at the lower energy level in the larger dot, optical excitation generated at QD_S goes back and forth in the resonant energy level (nutation), resulting in radiation from QD_S (Fig. 1.2b). The following section describes some of the emergent characteristics.

1.2.2 Space-, Time-, and Energy-Related Basic Functions

1.2.2.1 Global Summation Based on Optical Excitation Transfer: A Space-Domain Related Fundamental

The global summation, or data gathering, denoted by $\sum_{i=1}^{N} x_i$, where x_i represents N binary bits, is an important basic function in a wide range of electrical and optical devices and systems, such as optical code-division multiplexing (OCDM) [45], optical correlators [46], and content addressable memory (CAM) LSI chips [47, 48], among others. In known optical methods, wave propagation in free-space or in waveguides, using focusing lenses or fiber couplers, for example, is well-matched with such a data gathering scheme because the physical nature of propagating light is inherently suitable for the collection or distribution of information. However, the achievable level of integration of these methods is restricted due to the diffraction limit of light. In nanophotonics, on the other hand, the near-field interaction is inherently physically local.

The global data gathering mechanism, or summation, is realized based on the unidirectional energy flow via an optical near-field, as schematically shown in Fig. 1.3a, where surrounding excitations are transferred towards a quantum dot QD_C located at the center through optical near-field interactions [13, 49]. The lowest energy level in each quantum dot is coupled to a free photon bath to sweep out the excitation radiatively. The output signal is proportional to the lowest energy level in QD_C.

An experiment was performed to verify the nanoscale summation using CuCl quantum dots in an NaCl matrix, which has also been employed for demonstrating nanophotonic switches [50] and optical nano-fountains [13]. A quantum dot arrangement in which three small QDs (QD_1 to QD_3) surround a large QD at the center (QD_C) was chosen. Here, at most three light beams with different wavelengths, 325, 376, and 381.3 nm, are radiated to excite the respective quantum dots QD_1 to QD_3, having sizes of 1, 3.1, and 4.1 nm. The excited excitons are transferred to QD_C, and their radiation is observed by using a near-field fiber probe tip. Notice the output

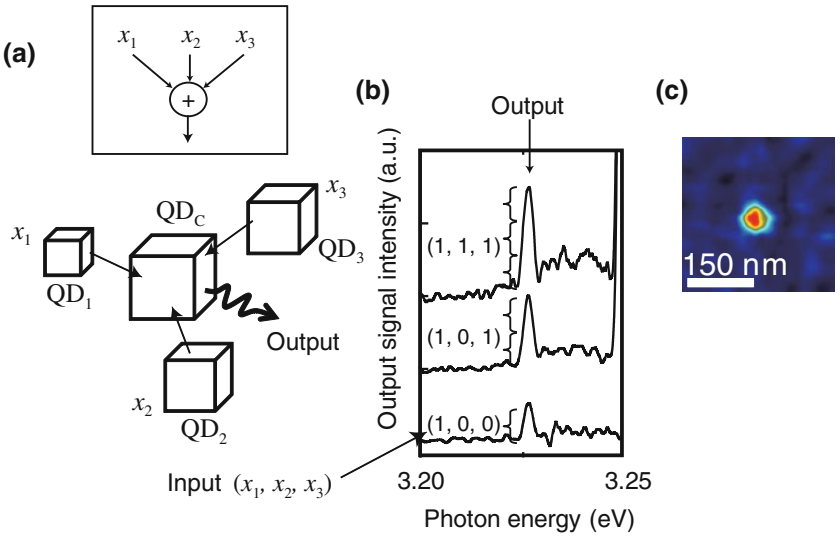


Fig. 1.3 **a** Architecture for global summation by optical excitation transfer. **b** Output signal intensity in frequency domain and **c** space-domain distribution

signal intensity at a photon energy level of 3.225 eV in Fig. 1.3b, which corresponds to a wavelength of 384 nm, or a QD_C size of 5.9 nm. The ratio of the intensities is approximately 1:2:3, depending on the number of excited QDs in the vicinity, as observed in Fig. 1.3b. The spatial intensity distribution was measured by scanning the fiber probe, as shown in Fig. 1.3c, where the energy is converged at the center. Hence, this architecture works as a summation mechanism that counts the number of input channels, based on exciton energy transfer via optical near-field interactions. Also, recently, stacked InAs quantum dots have been used to demonstrate room-temperature optical excitation transfer [51].

1.2.2.2 Optical Pulsation Based on Optical Excitation Transfer: A Time-Domain Related Fundamental

Generating an optical pulse train is one of the most important functionalities required for optical systems. Conventional principles of optical pulse generation are typically based on optical energy build-up in a cavity whose size is much larger than the optical wavelength; thus, the volume and the energy efficiency of the entire system have serious limitations. For nanophotonic applications, novel principles should be developed on the nanometer scale. In fact, Shojiguchi et al. theoretically investigated the possibility of generating superradiance in N two-level systems interacting with optical near-fields [52]. This approach, however, requires precise control of the initial states, which is not straightforward to implement. Here, we theoretically

demonstrate an optical pulsation method based on optical near-field interactions pumped by continuous-wave (CW) light irradiation [53]. With an architecture composed of two subsystems each of which involves excitation transfer based on optical near-field interactions, we observe pulsation in the populations based on a model system using a density matrix formalism. The details are described in [53].

As discussed in Sect. 1.2.1, when the lower level of QD_L is populated by an external input, the optical excitation occurring in QD_S cannot be transferred to QD_L because the lower energy level in QD_L is populated, which is called the state filling effect. Putting it another way, the population of the $(1, 0)$-level in QD_S is changed by the external input applied to L_1 in QD_L. Optical pulsation based on optical excitation transfer comes from the idea that the externally applied change induced in L_1 can be provided in a self-induced manner by S_1 with a certain timing delay. If QD_S is irradiated with continuous input light, such a change should repeat with a certain period; that is, a pulsed signal should result.

We consider two quantum dot systems, each of which consists of one smaller and one larger QD, as shown in Fig. 1.4a. One system, called *System* 1 hereafter, is represented by one smaller dot (QD_C) and one larger dot (QD_G). A CW input is provided to the upper level of QD_C. The optical near-field interaction between QD_C and QD_G is denoted by U_{CG}. Another system, called *System* 2, provides a delay time by multiple use of smaller and larger dots, as already experimentally realized in reference [54]. However, modeling the delay caused by multiple QDs makes the discussion of pulsation mechanisms unnecessarily complicated; thus, we assume an arbitrary delay time applied to the input signal of System 2, denoted by Δ, followed by the last two quantum dots in the delay system, namely, a smaller QD (QD_A) and a larger QD (QD_B), as indicated in Fig. 1.4a. Here, QD_A accepts radiation from QD_C in System 1; that is, the change of the states in QD_C is transferred to QD_A. The optical near-field interaction between QD_A and QD_B is denoted by U_{AB}. The output from QD_B then influences the lower energy level of QD_G.

We described the details of the above modeling based on a density matrix formalism. For example, regarding System 1, there are in total three energy levels (namely, C_1 in QD_C, and G_1 and G_2 in QD_G). The radiative relaxation rates from C_1 and G_1 are respectively given by γ_C and γ_G. Then we rigorously derived the quantum master equations for System 1 and System 2, respectively [53].

We assume the following typical parameter values based on experimental observations of energy transfer observed in ZnO quantum dots [55]: inter-dot optical near-field interactions (144 ps), sublevel relaxation (10 ps), and radiative decay times of the smaller dot (443 ps) and the large dot (190 ps). Also, we assume 1 ns for the delay Δ in System 2. Figure 1.4b demonstrates an example of the evolution of the populations involving the lower level of QD_G (G_1) for different CW input light amplitudes, where optical pulsation is successfully observed with appropriate input light amplitudes, as summarized by the peak-to-peak population of the pulsations shown in Fig. 1.4c. The detailed parameter dependencies were discussed in [53], validating the importance of optical excitation transfer.

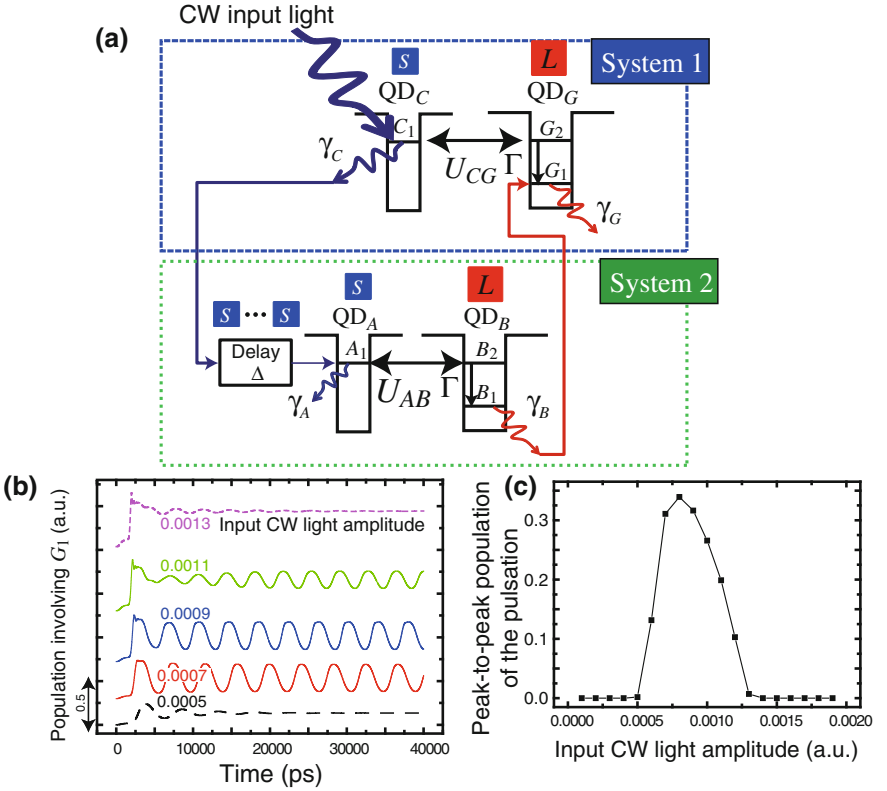


Fig. 1.4 **a** System architecture of pulsation composed of two subsystems each of which provides optical excitation transfer from QD$_S$ to QD$_L$. **b** Evolution of population in the lower energy level of QD$_G$ with different CW input light amplitudes. **c** Peak-to-peak population of pulsation versus input CW light amplitude

1.2.2.3 Energy Efficiency

We now discuss the lower bound of energy dissipation required for an optical excitation transfer by introducing two representative systems (System A and System B) [56]. The first one, System A in Fig. 1.5a, consists of two closely located quantum dots, and thus, optical excitation transfer from QD$_S$ to QD$_L$ occurs. We assume an interaction time U_{SL}^{-1} of 100 ps for System A, denoted by $U_{\mathrm{A}}^{-1} = 100$ ps in Fig. 1.5a. Such an interaction time is close to that of experimentally observed optical near-field interactions in CuCl QDs (130 ps) [50], ZnO quantum-well structures (130 ps) [20], ZnO QDs (144 ps) [55], and CdSe QDs (135 ps) [54]. The intersublevel relaxation time due to exciton–phonon coupling is in the 1–10 ps range [41, 42, 57], and here we assume $\Gamma^{-1} = 10$ ps. In System B on the other hand (Fig. 1.5a), the two quantum dots are intentionally located far away from each other. Therefore, the interactions between QD$_S$ and QD$_L$ should be negligible, and thus, optical excitation transfer

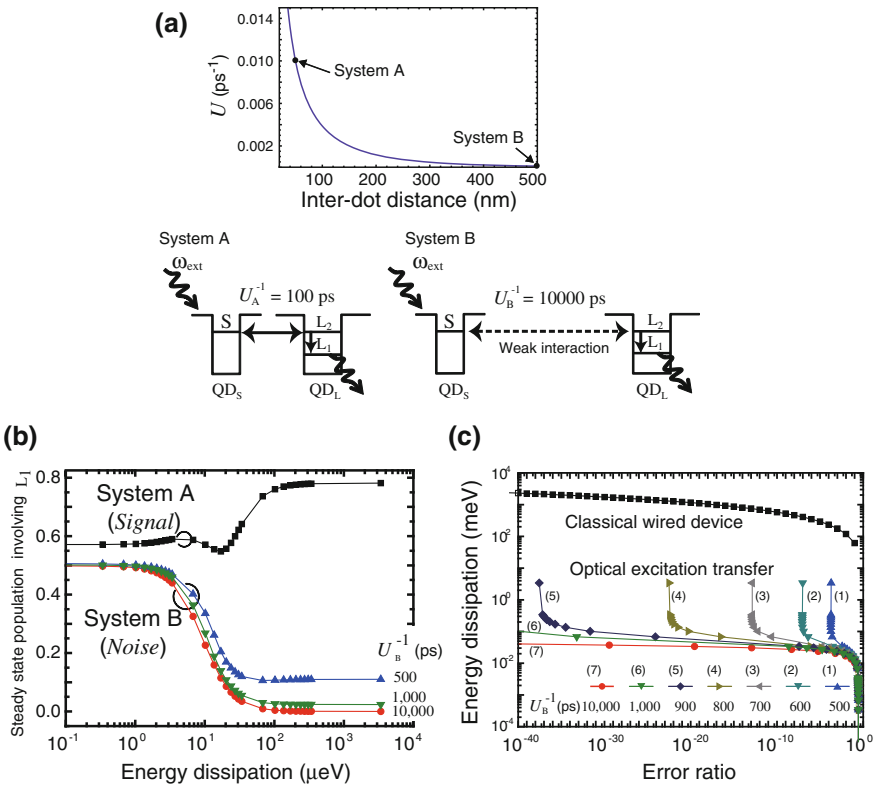


Fig. 1.5 Energy efficiency of optical excitation transfer. **a** Yukawa-type screened potential of an optical near-field interaction between two QDs as a function of the inter-dot distance We assume two systems consisting of a small QD (QD_S) and a large QD (QD_L): System A has a strong inter-dot near-field interactions (100 ps), whereas System B has negligible interaction (10,000 ps). **b** Steady-state population involving energy level L_1 in System A (squares) and System B as a function of the energy dissipation. For System B, three different cases are shown, with U_B^{-1} of 500, 1,000, and 10,000 ps, respectively indicated by, and marks. **c** Energy dissipation as a function of error ratio regarding optical excitation transfer and the bit flip energy required in a CMOS logic gate. The energy dissipation of optical excitation transfer is about 10^4-times lower than that in classical electrically wired devices

from QD_S to QD_L should not occur; namely, the radiation from QD_L should normally be zero. We assume $U_{SL}^{-1} = 10{,}000\,\text{ps}$ for System B, denoted by U_B^{-1} in Fig. 1.5a, indicating effectively no interactions between the two.

One remark here is that the inter-dot interaction times of System A and System B are related to the distances between the two quantum dots. The optical near-field interaction between two nanoparticles is known to be expressed as a screened potential using a Yukawa function, given by

$$U = \frac{A \exp(-\mu r)}{r}, \tag{1.6}$$

where r is the distance between the two [58]. In this representation, the optical near-field is localized around nanoparticles, and its decay length is equivalent to the particle size. Here, it should be noted that the inter-dot distance of System B indicates how close independent functional elements can be located. In other words, the interaction time of System B is correlated with the integration density of the total system. In order to analyze such spatial density dependencies, we assume that the U_{SL}^{-1} values of 100 and 10,000 ps respectively correspond to inter-dot distances of 50 and 500 nm. Here, the stronger interaction (100 ps) has been assumed, as already mentioned, based on a typical interaction time between closely spaced quantum dots. We also assume that the interaction with negligible magnitude (10,000 ps) corresponds to a situation where the inter-dot distance is close to the optical wavelength. Figure 1.5a shows the Yukawa-type potential curve given by (1.6).

When we assume a longer duration of the input light, the population converges to a steady state. When a pulse with a duration of 10 ns at the same wavelength (365 nm) is radiated, Fig. 1.5b summarizes the steady state output populations involving energy level L_1 evaluated at $t = 10$ ns as a function of the energy dissipation. The intended system behavior, that is, a higher output population in System A and a lower one in System B, is obtained in the region where the energy dissipation is larger than around 25 μeV. If we treat the population in System A as the amplitude of a "*signal*" and that in System B as "*noise*", the signal-to-noise ratio (SNR) can be evaluated based on the numerical values in Fig. 1.5b. To put it another way, from the viewpoint of the destination QD (or QD_L), the signal should come from a QD_S in its proximity (as in the case of System A), not from a QD_S far from QD_L (as in the case of System B); such a picture will aid in understanding the physical meaning of the SNR defined here. Also, here we assume that the input data are coded in an external system, and that QD_S is irradiated with input light at frequency ω_{ext}. With the SNR, the error ratio (P_E), or equivalently the bit error rate (BER), is derived by the formula $P_E = (1/2)erfc(\sqrt{SNR}/2\sqrt{2})$ where $erfc(x) = 2/\sqrt{\pi}\int_x^\infty \exp(-x^2)dx$, called the complementary error function [59]. The circles in Fig. 1.5c represent the energy dissipation as a function of the error ratio assuming the photon energy used in the above study (3.4 eV). According to [60], the minimum energy dissipation (E_d) in classical electrically wired devices (specifically, the energy dissipation required for a single bit flip in a CMOS logic gate) is given by $E_d = k_B T \ln\left(\sqrt{3}P_E/2\right)$, which is indicated by the squares in Fig. 1.5c. For example, when the error ratio is 10^{-6}, the minimum Δ in the optical excitation transfer is about 0.024 meV, whereas that of the classical electrical device is about 303 meV; the former is about 10^4 times more energy efficient than the latter.

As mentioned earlier, the performance of System B depends on the distance between the QDs. When the interaction time of System B (U_{B}^{-1}) increases, such as 500 ps, the steady state population involving L_1 is as indicated by the triangular marks in Fig. 1.5b; the population stays higher even with increasing energy dissipation

compared with the former case of $U_B^{-1} = 10{,}000$ ps. This means that the lower bound of the SNR results in a poorer value. In fact, as demonstrated by the triangular marks (1) in Fig. 1.5c, the BER cannot be smaller than around 10^{-4}, even with increasing energy dissipation. The lower bound of the BER decreases as the interaction time U_B^{-1} increases (namely, weaker inter-dot interaction), as demonstrated by the triangular and square marks (2) to (6) in Fig. 1.5c.

Finally, here we make a few remarks regarding the discussion above. First, we assumed arrays of identical independent circuits in the above discussion of density. Therefore, two circuits need spatial separations given by U_B so that unintended behavior does not occur. However, when two adjacent nanophotonic circuits are operated with different optical frequencies so that they can behave independently [61], those two circuits could be located closer together, which would greatly improve the integration density as a whole. Further analysis and design methodologies of complex nanophotonic systems, as well as comparison to electronic devices, is another topic to be pursued in future work. Second, because the energy separation in a single destination QD is limited by its size and lies in the range of meV, the results for energy separations in the μeV range correspond to cases where the destination dot QD_L represents a theoretical model of a coupled quantum dot system such as a pair of quantum dots. The coupled system exhibits optical near-field interactions with the smaller QD, followed by inter-dot electron transfer resulting in optical radiation. In fact, Matsumoto et al. have demonstrated spin-dependent carrier transfer leading to optical radiation between a coupled double quantum well system composed of magnetic and nonmagnetic semiconductors [62], which can be applied to quantum dot systems [63]. Third, a discussion of input and output interfaces is necessary. The above discussion has focused on the lower bound of energy dissipation in the quantum dots. Practical operation of real devices requires input and output interfaces, and the minimum number of photons for a bit slot may be of concern when taking account of noise at the receivers. In [64], Naruse et al. unified such considerations into an evaluation model and analyzed experimental results based on stacked QDs [19]. It was found that optical excitation transfer still exhibits around 10^4-times better energy efficiency compared with electronic counterparts [64].

1.2.3 Going Beyond the von Neumann Architecture: Stochastic Solution Searching

Another aspect of the stochastic nature inherent in nanophotonics is its application to novel computing devices and architectures [65]. Nature-inspired architectures are attracting significant attention in various research arenas, such computational neurosciences, stochastic-based computing and noise-based logic, and spatiotemporal computation dynamics [66], in order to benefit from the superior attributes of nature and living systems.

Among this research, Aono et al. demonstrated "amoeba-based computing" tasks, such as solving the constraint satisfaction problem (CSP) [66] and the traveling

salesman problem (TSP) [67], by utilizing the spatio-temporal oscillatory dynamics of the photoresponsive amoeboid organism *Physarum* combined with external optical feedback control. These demonstrations indicate that spatiotemporal stochastic dynamics can be utilized for obtaining solutions for problems which today's von Neumann-architecture computers cannot deal with efficiently. In particular, it should be noted that the optical excitation transfer between quantum nanostructures mediated by optical near-field interactions is fundamentally probabilistic, as indicated by the quantum master equations. Until energy dissipation is induced, an optical excitation simultaneously interacts with all potentially transferable destination quantum dots in the resonant energy level. Such a probabilistic behavior can be used for solution searching and exploration. In addition, the optical energy transfer has been shown to be 10^4-times more energy efficient than that of the bit-flip energy required in conventional electrically wired devices, as discussed in Sect. 1.2.2.3 [56].

Here, we investigate the spatiotemporal dynamics inherent in optical excitation transfer. Furthermore, we demonstrate that it can be utilized for solving a CSP. The optical excitation transfer depends on the existence of resonant energy levels between the quantum dots (QDs) or the state filling effect occurring at the destination QDs. Such a spatial and temporal mechanism yields different evolutions of energy transfer patterns combined with certain feedback mechanisms. In contrast to biological substrates, optical energy transfer is implemented by highly-controlled engineering means for designated structures. The operating speed of such optical near field-mediated QD systems, which is on order of nanoseconds when we are concerned with radiative relaxation processes, is also significantly faster than ones based on biological organisms, which is on the order of seconds or minutes [66].

In addition, we should emphasize that the concept and the principles discussed here are fundamentally different from those of conventional optical computing or optical signal processing, which are limited by the abilities of propagating light. The concept and the principles are also different from the quantum computing paradigm where the superposition of all possible states is exploited to lead to a correct solution. The optical near field-mediated energy transfer is a coherent process, indicating that an optical excitation could be transferred to all possible destination QDs via a resonant energy level, but such coherent interaction between QDs results in a unidirectional energy transfer by means of an energy dissipation process occurring in the larger dot. Thus, our approach paves the way to another computing paradigm in which both coherent and dissipative processes are exploited.

Here we assume one smaller quantum dot, denoted by QD_S, and four larger quantum dots, denoted by QD_{L1}, QD_{L2}, QD_{L3}, and QD_{L4}, as shown in Fig. 1.6a. The smaller and larger QDs are resonant with each other. Figure 1.6b shows representative parameterizations associated with the system; for example, the (1, 0)-level in the smaller QD is denoted by S, and the (1,1)-level in QD_{L_i} is denoted by $\mathrm{L_i^{(U)}}$. These levels are resonant with each other and are connected by inter-dot interactions denoted by U_{SL_i} ($i = 1, \ldots, 4$). The lower level in QD_{L_i}, namely, the (1, 0)-level, is denoted by $\mathrm{L_i^{(L)}}$, which could be filled via the sublevel relaxation from $\mathrm{L_i^{(U)}}$ denoted by Γ_{L_i}. The radiations from the S and L_i levels are respectively represented by the relaxation

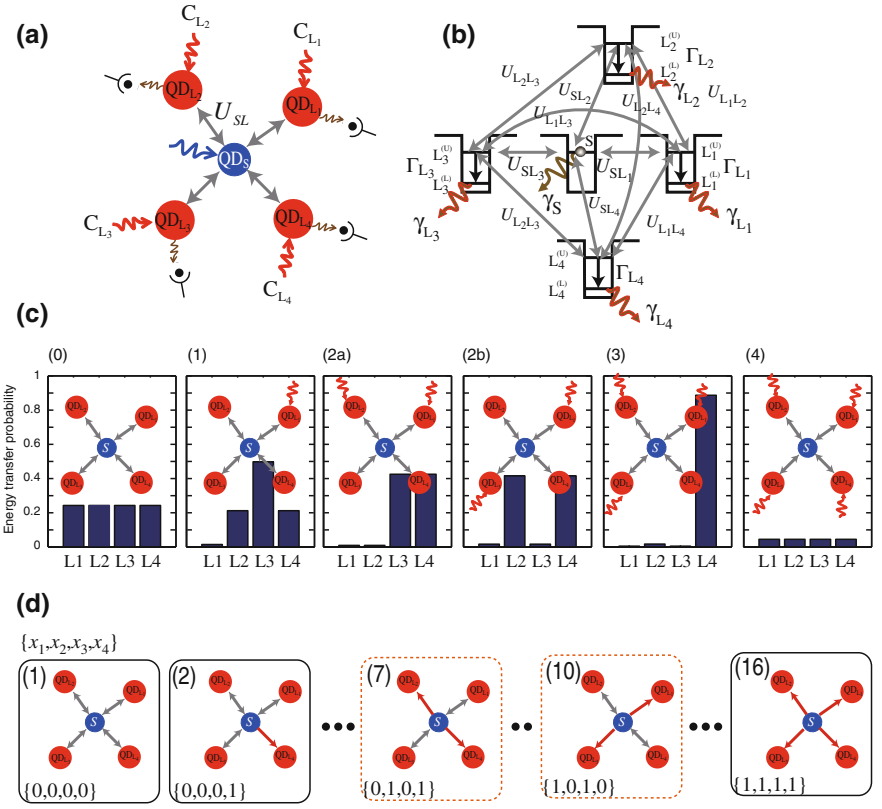


Fig. 1.6 **a** Architecture of the optical-energy-transfer–based system for solving constraint satisfaction problem and composed of a smaller quantum dot and four larger quantum dots **b** and its energy diagram. Radiation from the larger quantum dots is detected. Control light is used for inducing state filling in the larger quantum dots. **c** Energy transfer probabilities calculated as time integrals of the populations depending on the state filling of the larger dots, shown in (**a**). There are a total of six different spatial arrangements of the induced state filling, excluding the symmetries. **d** There are a total of 2^4 (=16) different combinations of binary values in the given problem. The correct solutions are the state numbers (7) and (10), where $\{x_1, x_2, x_3, x_4\}$ are respectively given by {0, 1, 0, 1} and {1, 0, 1, 0}

constants γ_S and γ_{Li}. We call the inverse of those relaxation constants the radiation lifetime in the following. We also assume that the photon radiated from the lower level of QD_{L_i} can be separately captured by photodetectors. The channels of control light, denoted by C_{L_i}, can induce a state filling effect at $L_i^{(L)}$. Summing up, Fig. 1.6a schematically represents the architecture of the system to be studied in this section for solving a CSP.

In the numerical calculation, we assume $U_{SLi}^{-1} = 100\,\text{ps}$, $\Gamma_i^{-1} = 10\,\text{ps}$, $\gamma_{Li}^{-1} = 1\,\text{ns}$, and $\gamma_S^{-1} = 2.92\,\text{ns}$ as a typical parameter set. If there is no state filling in the system, an optical excitation sitting initially at S can be transferred to any one of QD_{L1} to

QD_{L4} with the same probability, as demonstrated in Fig. 1.6c, 0, which is the time-integral of the population involving the energy level $L_i^{(L)}$. If QD_{L1} suffers from state filling, on the other hand, the initial excitation at S is more likely to be transferred to QD_{L2}, QD_{L3}, or QD_{L4}, as shown in Fig. 1.6c, 1. Looking at the results more closely, the probability of transfer to QD_{L3} is higher than the probabilities of transfer to QD_{L2} and QD_{L4} by considering the geometrical arrangements of the system. A detailed discussion is found in [65]. Additionally, the energy transfer probabilities in the presence of two-, three-, or four-state filling are summarized in Fig. 1.6c. The energy transfer probability, given by the integral of the population divided by a constant gain factor, is a figure-of-merit (FoM) indicating the trend of optical energy transfer from the smaller quantum dot to the four larger ones. This energy transfer probability does *not* obey the conservation law of probability, namely, the summation of the transition probabilities to QD_{L_i} is not unity. Instead, we see that the energy transfer to QD_{L_i} occurs if a random number generated uniformly between 0 and 1 is less than the transition probability to QD_{L_i}.

The idea for problem solving is to control optical energy transfer by controlling the destination QD by using control light in an adequate feedback mechanism. We assume that photon radiation, or observation, from the energy level $L_i^{(L)}$ is equivalent to a binary value x_i resulting in a logical 1 level, whereas no observation of a photon means $x_i = 0$.

We consider the following constraint satisfaction problem as an example regarding an array of N binary-valued variables $x_i (i = 1, \ldots, N)$. The constraint is that $x_i = NOR(x_{i-1}, x_{i+1})$ should be satisfied for all i. That is, variable x_i should be consistent with a logical NOR operation of the two neighbors. For $i = 0$ and N, the constraints are respectively given by $x_1 = NOR(x_N, x_2)$ and $x_N = NOR(x_{N-1}, x_1)$. We call this problem the "NOR problem" in the following. Taking account of the nature of an individual NOR logic operation, one important inherent character is that if $x_i = 1$ then its two neighbors should both be zero, or $x_{i-1} = x_{i+1} = 0$. Recall that a photon radiated, or observed, from the energy level $L_i^{(L)}$ corresponds to a binary value $x_i = 1$, whereas the absence of an observed photon means $x_i = 0$. Therefore, $x_i = 1$ should mean that the optical energy transfer to both $L_{i-1}^{(L)}$ and $L_{i+1}^{(L)}$ is prohibited so that $x_{i-1} = x_{i+1} = 0$ is satisfied. Therefore, the feedback or control mechanism is as follows:

[Control mechanism] If $x_i = 1$ at cycle t, then the control light beams C_{i-1} and C_{i+1} are turned on at cycle $t = t+1$.

In the case of $N = 4$, there are in total 2^4 optical energy transfer patterns from the smaller dot to larger ones. In this case, variables satisfying the constraints do exist, and they are given by $\{x_1, x_2, x_3, x_4\} = \{0, 1, 0, 1\}$ and $\{1, 0, 1, 0\}$, which we call "correct solutions". Figure 1.6d schematically represent some of the possible states, where the states (7) and (10) respectively correspond to the correct solutions.

We now make a few remarks regarding the NOR problem. One is about potential deadlock, analogous to Dijkstra's "*dining philosophers problem*", as already argued by Aono et al. in reference [66]. Starting with an initial state $x_i = 0$ for all i, and assuming a situation where optical energy is transferred to all larger QDs, we observe

photon radiation from all energy levels $L_i^{(L)}$, namely, $x_i = 1$ for all i. Then, based on the feedback mechanism shown above, all control light beams are turned on. If such a feedback mechanism perfectly inhibits the optical energy transfer from the smaller QD to the larger ones at the next step $t + 1$, the variables then become $x_i = 0$ for all i. This leads to all control light beams being turned off at $t + 2$. In this manner, all variables constantly repeat a periodic switching between $x_i = 0$ and $x_i = 1$ in a synchronized manner. Consequently, the system can never reach the correct solutions. However, as indicated in Fig. 1.6c, the probability of optical energy transfer to larger dots is in fact not zero even when all larger QDs are illuminated by control light, as shown in Fig. 1.6c, 4. Also, even for a non-illuminated destination QD, the energy transfer probability may not be exactly unity. Such a stochastic behavior of the optical energy transfer is a key role in solving the NOR problem. This nature is similar to the demonstrations in amoeba-based computing [66] where fluctuations of chaotic oscillatory behavior involving spontaneous symmetry breaking in the amoeboid organism guarantee such a critical property.

The operating dynamics cause one pattern to change to another one in every iteration cycle. Thanks to the stochastic nature, each trial could exhibit a different evolution of the energy transfer patterns. In particular, the transition probability, shown in Fig. 1.6c, affects the behavior of the transitions.

The curves in Fig. 1.7a represent the evolution of the output appearance from QD_{L_i}, namely, the ratio of the incidence when $x_i = 1$ among 1,000 trials evaluated at each cycle. The curves in Fig. 1.7b characterize the ratio of the appearance of the state that corresponds to the correct solutions: {0, 1, 0, 1} (state 7) and {1, 0, 1, 0} (state 10). When we closely examine the evolutions of x_i in Fig. 1.7a, we can see that the pair x_1 and x_3 exhibit similar behavior, as do the pair x_2 and x_4. Also, the former pair exhibit larger values, whereas the latter pair exhibit smaller values, and vice versa. This corresponds to the fact that correct solutions are likely to be reached as the iteration cycle increases.

Such a tendency is more clearly represented when we evaluate the time-averages of the characteristics in Fig. 1.7a, b. Figure 1.7c shows the evolutions of the ratio of the incidences when $x_i = 1$, and Fig. 1.7d shows the ratios of State (7) and State (10) averaged over every 5 cycles. We can clearly observe a similar tendency to the one described above. Also, we should emphasize that, thanks to the probabilistic nature of the system, the states of correct solutions appear in an interchangeable manner. This is a clear indication of the fact that the probabilistic nature of the system autonomously seeks the solutions that satisfy the constraints of the NOR problem; the state-dependent probability of energy transfer plays a critical role in this. In other words, it should be emphasized that a non-local correlation is manifested in the evolution of x_i; for instance, when the system is in State (7), {0, 1, 0, 1}, the probabilities of energy transfer to QD_{L1} and QD_{L3} are equally comparably low (due to state filling), whereas those to QD_{L2} and QD_{L4} are equally comparably high, indicating that the probability of the energy transfer to an individual QD_{L_i} has inherent spatial patterns or non-local correlations. At the same time, the energy transfer to each QD_{L_i} is indeed probabilistic; therefore, the energy transfer probability to, for

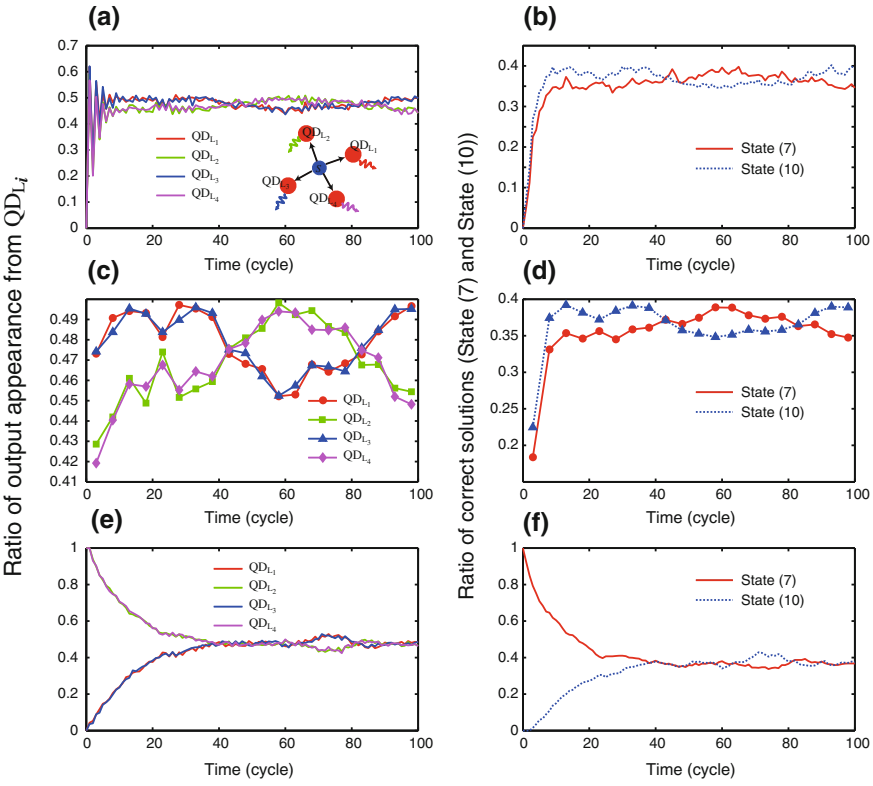


Fig. 1.7 **a** The evolution of the ratio of the output appearance from $\mathrm{QD}_{\mathrm{L}_i}$, or $x_i = 1$, and **b** the ratios of the states corresponding to the correct solutions with the initial state of $\{x_1, x_2, x_3, x_4\} = \{0, 0, 0, 0\}$. **c**, **d** Time-averaged traces of (**a**) and (**b**), respectively. **e** The evolution of the ratio of the output appearance from $\mathrm{QD}_{\mathrm{L}_i}$, or $x_i = 1$, and **f** the ratios of the states corresponding to correct solutions with the initial state $\{x_1, x_2, x_3, x_4\} = \{0, 1, 0, 1\}$

instance, QD_{L1} is not zero even in State (7), and thus, the state could transition from State (7) to State (10), and vice versa. In fact, starting with the initial condition of State (7), the ratio of the output appearance from QD_{L1} and the ratio of the correct solutions evolve as shown in Fig. 1.7e, f, where States (7) and (10) occur equally in the steady state at around 20 time cycles.

We make two final remarks to conclude this section. The first is about the relevance to a satisfiability problem (SAT). In the case of $N = 4$, solving the NOR problem demonstrated above is equivalent to solving the following satisfiability problem instance given in a conjunctive normal form:

$$f(x_1, x_2, x_3, x_4) = (\neg x_1 \vee \neg x_2) \wedge (\neg x_1 \vee \neg x_4) \wedge (\neg x_2 \vee \neg x_3) \wedge (\neg x_3 \vee \neg x_4) \wedge (x_1 \vee x_2 \vee x_3) \wedge (x_1 \vee x_2 \vee x_4) \wedge (x_1 \vee x_3 \vee x_4) \wedge (x_2 \vee x_3 \vee x_4). \quad (1.7)$$

Since the maximum number of literals in clauses in (1.7) is three, this is an instance of a so-called 3SAT problem [68]. We have already reported that such a SAT problem could be dealt with by variants of our optical-near-field–mediated systems [69]. SAT is an important nondeterministic polynomial-time complete (NP-complete) problem, indicating that no fast algorithm has been found yet [68]. We consider that nanophotonic principles could potentially provide a new way to solve such computationally demanding problems. In addition, Kim et al. have succeed in applying optical excitation transfer to decision making applications, giving better performance than conventional approaches [70].

The second remark is about the implementation of optical energy transfer for such stochastic computing applications. As mentioned in the introduction, rapid advancements have been made recently in nanomaterials for optical energy transfer [19, 20, 26, 51, 57]. Among various technologies, for example, Akahane et al. successfully demonstrated energy transfer in multi-stacked InAs QDs, where layer-by-layer QD size control has been accomplished [51]. Adequate QD size control also allows optical coupling between optical far-fields and optically-allowed energy levels in a quantum dot mixture, which could help to solve the interfacing issues of the system.

1.3 Nanophotonics for Security

The security aspects of optics have been studied extensively [4], and some of them have been commercialized, such as in optical document security [71]. However, since all of the existing optical security principles and technologies are based on optical far-fields, or propagating light, such as Fourier optics [46], they suffer from associated limitations, such as the difficulty of miniaturizing devices and systems beyond dimensions limited by the diffraction of light [8], leading also to the difficulty in applying them to tamper-resistant hardware in embedded systems etc.

On the other hand, nanophotonics, which utilizes light and matter interactions at scales below the wavelength of light, has recently attracted attention [11]. In addition to the ability to break through the diffraction limit of light, we can also make use of the unique physical processes in nanophotonics, such as optical energy transfer via optical near-field interactions, as discussed in Sect. 1.2. For example, from a security standpoint, the tamper resistance of optical excitation transfer has been demonstrated [72].

At the same time, for tamper-resistant hardware, we can make use of electromagnetic fields in the subwavelength regime, in other words, near-field optics or nano-optics [8]. In particular, technologies allowing shape-engineering of nanostructures,

such as electron beam lithography, are stable and reliable technologies for implementing systems and devices. We have demonstrated a "*hierarchical hologram*" that works in both optical far-fields and near-fields, the former being associated with conventional holographic images, and the latter being associated with the optical intensity distribution originating from a nanometric structure embedded in the hologram, which is accessible only via optical near-fields [22, 73, 74]. In other words, information hiding can be realized by using optical near-fields and nanofabrication technologies. Also, authentication functions can be implemented by using two shape-engineered nanostructures and their associated optical near-fields [21, 75]. In this system, the two nanostructures respectively work as a *lock* and *key*, where authenticity is guaranteed by the nanoscale-precision shapes of the structures.

As described above, the physical principles of nano-optics will contribute to novel security means. However, from a security perspective, more-solid theoretical foundations and performance evaluations are necessary. In this section, we present a fundamental theory of nano-optics based on a rigorous treatment and evaluate its associated performance. We then characterize particular examples of nano-optical security means based on this theoretical foundation. These investigations will lead to enriched tamper-resistant hardware utilizing nano-optical processes, such as physical unclonable functions (PUFs) or security technologies in embedded systems in general.

1.3.1 Theoretical Foundation

Optical near-fields are the localized, non-propagating components of electromagnetic fields in the vicinity of materials [8]. We need to locate certain kinds of *reader* to induce interactions between the material under study and the reader. In order to characterize the structure of the system, we denote the entities of the system as follows. Let the material under study, or the device, be denoted by D, and the reader by R. The output signal is written as $v = g(D, R)$.

One of the characteristic consequences of nano-optical systems is that the output signal depends sensitively on both D and R, which is represented by the function $g(D, R)$. Such physically inherent properties of nano-optics are well-matched with the architecture of tamper-resistant hardware. In order to theoretically represent the fundamental characters, we describe the system as follows. The device D is represented by a point dipole located a distance X away from the origin, and the reader R is located a distance Z away from the origin, as shown in Fig. 1.8a. The R-dependence of the output signal is regarded as being equivalent to the Z-dependence.

Here, we employ the angular spectrum representation of electromagnetic fields, in which the electromagnetic fields are represented as a superposition of evanescent waves with different decay lengths and corresponding spatial frequencies [76–78]. As well as giving a rigorous analytical treatment of near-field components in the subwavelength regime, this picture can explicitly include the D- and Z-dependences as follows.

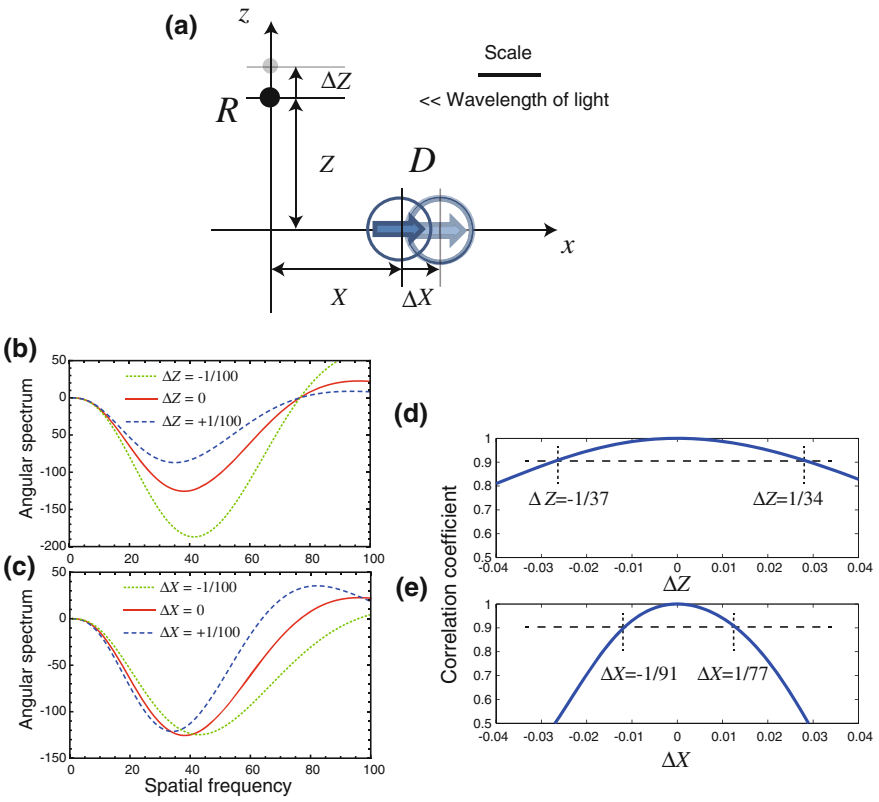


Fig. 1.8 Fundamental system model for nano-optical security. **a** A point dipole (D) and an evaluation point (Z) represent the device under study (D) and the reader (R), respectively. **b**, **c** The output signal is evaluated by the angular spectrum, which is the near-field component of the electromagnetic field in the subwavelength regime. Differences in Z and X are respectively shown in (**b**) and (**c**). **d**, **e** Correlation coefficient of the output signal as a function of minute differences in Z and X. A tiny difference strongly affects the output signal, which is a manifestation of the ability of nano-optics for tamper-resistant hardware, offering functions like anticounterfeiting, authentication, etc.

Suppose that there is an electric dipole, $\mathbf{D} = d(\cos\varphi, 0)$, on the xz plane oscillating at frequency K, as shown in Fig. 1.8a. The velocity of light is taken as unity. Now, consider the electric field of radiation observed at a position displaced from the dipole by $\mathbf{R} = (r_{||}\cos\phi, Z)$. The angular spectrum representation of the z-component of the optical near-field is given by

$$E_z(\mathbf{R}) = \left(\frac{iK^3}{4\pi\varepsilon_0}\right)\int_1^\infty ds_{||}\frac{s_{||}}{s_z}f_z(s_{||}, \mathbf{D}, \mathbf{R}), \tag{1.8}$$

where

$$f_z(s_{||}, \mathbf{D}, \mathbf{R}) = ds_{||}\sqrt{s_{||}^2 - 1}\cos(\varphi - \phi)J_1\left(Kr_{||}s_{||}\right)\exp\left(-KZ\sqrt{s_{||}^2 - 1}\right). \quad (1.9)$$

Here, $s_{||}$ is the spatial frequency of an evanescent wave propagating parallel to the x axis, and $J_n(x)$ represents Bessel functions of the first kind. The term $f_z(s_{||}, \mathbf{D}, \mathbf{R})$ is called the angular spectrum of the electric field. We consider that $f_z(s_{||}, \mathbf{D}, \mathbf{R})$ is equivalent to the signal $v = g(D, R)$ characterized in the system model. In the case of the system model shown in Fig. 1.8a, the parameters are given by $\varphi = 0$, $r_{||} = X$, and $\phi = \pi$. Note that X and Z can be effectively considered to be given in units of wavelength.

The solid curve in Fig. 1.8b shows the angular spectrum when $X = 1/20$ and $Z = 1/20$. This corresponds to an *authentic* device D and an *authentic* reader R. Differences of the reader R are equivalent to differences of Z; for instance, when Z is shifted by distance $\Delta Z = -1/100$ the angular spectrum is given by the dotted curve in Fig. 1.8b. Similarly, when $\Delta Z = 1/100$, the angular spectrum is given by the dashed curve in Fig. 1.8b. As shown by the changes of the curve in Fig. 1.8b, the slight difference with respect to Z results in a different output signal from the system. In order to quantitatively evaluate the Z-dependence, the correlation coefficient of the angular spectrum is calculated as a function of ΔZ as summarized in Fig. 1.8d. If we determine that an authentic signal should yield a correlation coefficient larger than 0.9, ΔZ should be between $-1/37$ and $1/34$, which would be an extremely small absolute value in real dimensions. This indicates that nano-optics provides an evident reader-dependence. Similarly, by considering the position of the dipole as the identity of the device, a different position of the dipole provides a different angular spectrum. The solid, dotted, and dashed curves in Fig. 1.8c respectively indicate the angular spectra when ΔX is given by 0, $-1/100$, and $+1/100$. The correlation coefficient is evaluated as shown in Fig. 1.8e; it is larger than 0.9 when ΔX is between $-1/77$ and $1/91$, indicating that the output signal is sensitive to subtle differences of the device D.

1.3.2 Hierarchical Hologram for Information Hiding and its Theoretical Fundamentals

Holography, which generates natural three-dimensional images, is one of the most common anti-counterfeiting techniques [71]. In a volume hologram, the surface is ingeniously formed into microscopic periodic structures that diffract incident light in specific directions. Generally, these microscopic structures are recognized as being difficult to duplicate; therefore, holograms have been widely used in the anti-counterfeiting of bank notes, credit cards, etc. However, conventional anti-counterfeiting methods based on the physical appearance of holograms are nowadays not completely secure [79]. Nano-optical solutions would provide higher

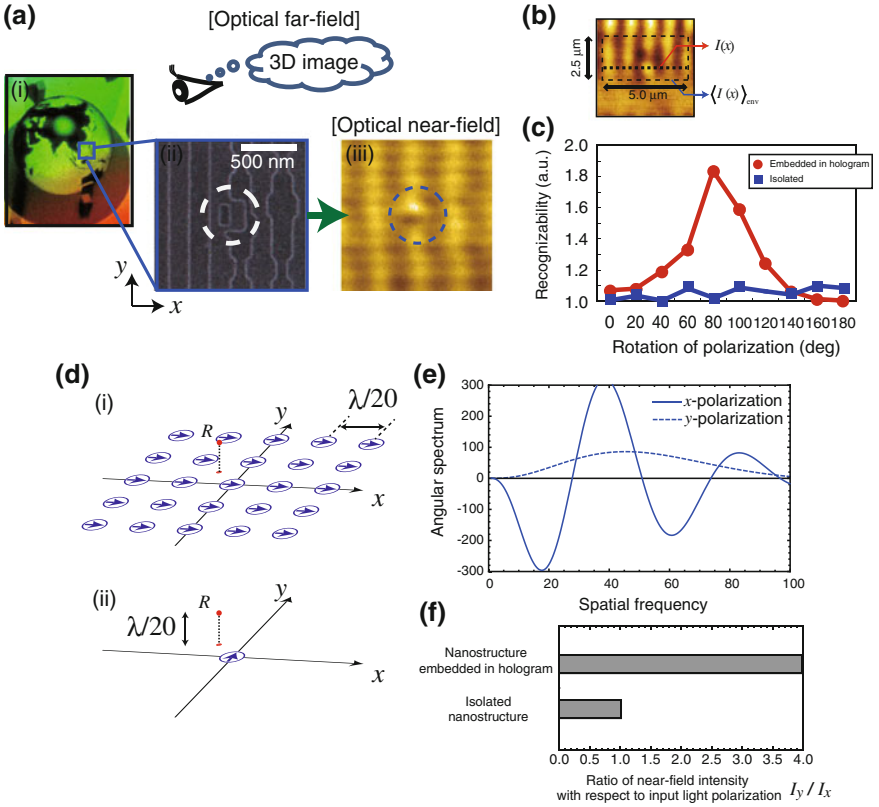


Fig. 1.9 Hierarchical hologram. Experimental demonstrations and theoretical foundations. **a** Hierarchical hologram. **b**, **c** Polarization dependencies in the near-field retrieval from the nanostructure embedded in the hologram. **d**, **e**, **f** Theoretical reasoning of the polarization dependencies of nanostructures embedded in holograms. **d**, **e** x-polarized light induces an array of dipoles, which results in weaker localization in its near-fields ((**d**, i) and solid curve in (**e**)), whereas y-polarized light induces dipoles only at the nanostructures embedded in the hologram, which results in high localization in its near-fields ((**d**, ii) and dashed curve in (**e**)). **f** These results explain the evident polarization dependencies

anti-counterfeiting capability and could potentially enable other applications, such as artifact-metric systems [80].

A *hierarchical hologram* works in both optical far-fields and near-fields, the former being associated with conventional holographic images (Fig. 1.9a, i), and the latter being associated with the optical intensity distribution originating from a nanometric structure (Fig. 1.9a, ii) that is accessible only via optical near-fields (Fig. 1.9a, iii). In principle, a structural change occurring at the subwavelength scale does not affect the optical response function, which is dominated by propagating light. Therefore, the visual aspect of the hologram is not affected by such a small structural change on the surface. Additional data can thus be written by engineering

structural changes in the subwavelength regime so that they are only accessible via optical near-field interactions, without having any influence on the optical response obtained via the conventional far-field light. As indicated in Fig. 1.9a, i, we can observe a three-dimensional image of the earth reconstructed from the device. More specifically, the device was based on the design of *Virtuagram*®, developed by Dai Nippon Printing Co., Ltd., Japan, which is a high-definition computer-generated hologram composed of binary-level one-dimensional modulated gratings, as shown in the scanning electron microscope (SEM) image in Fig. 1.9a, ii. Within the device, we slightly modified the shape of the original structure of the hologram so that the nanostructural change was accessible only via optical near-field interactions. As shown in Fig. 1.9a, ii, square- or rectangle-shaped structures, whose associated optical near-fields correspond to the additional or hidden information, were embedded in the original hologram structures. The unit size of the nanostructures ranged from 40 to 160 nm.

The original hologram structure is basically composed of one-dimensional gratings; that is, the structure is topologically connected along the vertical direction. The embedded nanostructure for the hidden information destroys such a connected topology of the original gratings.

From a security standpoint, we should mention two points. One is that counterfeiting such an intricate nanostructure would be technologically very difficult, if not impossible, since realizing the minimum feature size—40 nm in the device shown in Fig. 1.9a [22]—would require the attackers to have high-quality nanofabrication facilities. The other is that such a *topologically disconnected* structure exhibits strong polarization dependence. The input light induces oscillating surface charge distributions due to the coupling between the light and electrons in the metal. Note that the original 1D grid structures span along the vertical direction. The y-polarized input light induces surface charges along the vertical grids. Since the grid structure continuously exists along the y-direction, there is no chance for the charges to be concentrated. However, in the area of the embedded square-shaped nanostructure, we can find structural discontinuity in the grid; this results in higher charge concentrations at the edges of the embedded nanostructure. On the other hand, the x-polarized input light sees structural discontinuity along the horizontal direction due to the vertical grid structures, as well as in the areas of the embedded nanostructures. It turns out that charge concentration occurs not only in the edges of the embedded nanostructures but also at other horizontal edges of the environmental grid structures. When a square-shaped nanostructure is *isolated* in a uniform plane, both x- and y-polarized input light have equal effects on the nanostructures. The nanostructures embedded in holograms could exploit these polarization dependences.

In the experimental demonstration, near-field intensity distributions were detected using a near-field optical microscope operated in an illumination-collection mode with an optical fiber tip having a radius of curvature of 5 nm. The observation distance between the tip of the probe and the sample device was set at less than 50 nm. The light source used was a laser diode (LD) with an operating wavelength of 785 nm, and scattered light was detected by a photomultiplier tube (PMT). We examined near-field images in the vicinity of nanostructures that were embedded in the hologram and

nanostructures that were not embedded in the hologram using a linearly polarized radiation source with its polarization rotated from 0° to 180° at 20° intervals. In the case of nanostructures embedded in the hologram, clear polarization dependence was observed. To quantitatively evaluate the polarization dependence, we adopted a figure-of-merit, what we call *recognizability*, for the observed near-field images, which represents the difference of the near-field intensity compared with that of neighboring areas. More specifically, let the horizontal intensity profile along the dashed line in Fig. 1.9b, which crosses the area of the nanostructure, be denoted by $I(x)$, where $x$represents the horizontal position. Also, let the average intensity along the vertical direction at the horizontal position x within a range of 2.5 μm be denoted by $\langle I(x)\rangle_{\mathrm{env}}$, which indicates the environmental signal distribution. When a higher intensity is obtained selectively from the area of the nanostructure, the difference between $I(x)$ and $\langle I(x)\rangle_{\mathrm{env}}$ should be large. On the other hand, if the intensity distribution is uniform along the vertical direction, the difference between $I(x)$ and $\langle I(x)\rangle_{\mathrm{env}}$ should be small. Thus, the difference between $I(x)$ and $\langle I(x)\rangle_{\mathrm{env}}$ indicates the *recognizability* of the nanostructures. With respect to the polarization angle θ, we evaluate the *recognizability* $R(\theta)$ as $R(\theta) = \sum_x |I(x) - \langle I(x)\rangle_{\mathrm{env}}|$. The square and circular marks in Fig. 1.9c respectively show $R(\theta)$ based on the near-field images of isolated nanostructures and those embedded in the hologram. Clear polarization dependency is observed in the case of the nanostructures embedded in the hologram.

Here, we theoretically deal with the above properties of a hierarchical hologram using the theoretical elements presented in Sect. 1.3.1. As mentioned earlier, x-polarized input light sees structural discontinuity along the horizontal direction due to the vertical grid structures, as well as in the areas of the embedded nanostructures. We represent such a situation by a 5×5 array of dipoles arranged in the xy plane, as shown in Fig. 1.9d, and assume that the dipole located at the center corresponds to the nanostructure implemented for the information hidden within the hologram. We consider the electromagnetic field in the vicinity of the embedded nanostructure at the position R in Fig. 1.9d, i.

The angular spectrum is calculated by summing up all contributions from the 5×5 dipoles at the point R, as shown by the solid curve in Fig. 1.9e, which oscillates as a function of spatial frequency, indicating that the field is not strongly localized at point R. On the other hand, when the y-polarized input light is irradiated, the charge concentration occurs only at the embedded nanostructure, which is represented by Fig. 1.9d, ii. The corresponding angular spectrum at point R is given by the dashed curve in Fig. 1.9e. It exhibits a peak value at a certain spatial frequency, meaning that the higher intensity electric field is localized at point R. Such a difference is one theoretically predicted behavior that is observed experimentally. By calculating the integrals of the angular spectrum with respect to x- and y-polarized input light, which respectively represent the near-field intensities given by I_x and I_y, the ratio of the two, I_y/I_x, results in a value of about 4. In the case of isolated nanostructures, since both x- and y-polarized light components induce the same responses, and the corresponding ratio is unity, as summarized in Fig. 1.9f, these theoretical investigations are consistent with the experimental observations shown in Fig. 1.9c.

1.3.3 Shape-Engineered Nanostructures for Authentication Functions

In this section, we demonstrate that two metal nanostructures, called *Shape D* and *Shape R*, can be designed to exhibit far-field radiation only when their shapes are appropriately configured and when they are closely stacked [21, 75]. Such functionality is useful in ensuring authentication or certification, where a system should work only when the two nanostructures match, just like a *lock* and *key*. We explain the operating principle by observing the induced electric currents and their associated optical near-fields based on the theoretical elements presented in Sect. 1.3.1.

We have previously proposed a theoretical scheme that is useful for examining the relation between the shapes of nanostructures and the resulting light polarization in their optical near-field and far-field [81]. In that study, the geometry of a given nanostructure can be understood in terms of two factors, "individual element" and "layout", where the former represents the shapes of individual elements, and the latter their spatial arrangement. There are two important physical aspects in analyzing their corresponding optical responses. One is that input light induces electric currents within individual elements of the metal nanostructure. The other is optical near-fields between those individual elements. We can understand those two processes in a unified manner as vectorial elements in the system [81].

Shape D and Shape R were designed as rectangular units aligned on an xy-plane at constant intervals horizontally (along the x-axis) and vertically (along the y-axis), as respectively shown in Fig. 1.10a, b. When we irradiate Shape D with x-polarized light, surface charges are concentrated at the horizontal edges of each of the rectangular units. The relative phase difference of the oscillating charges between the horizontal edges is π, which is schematically represented by $+$ and $-$ marks in Fig. 1.10a. Now, consider the y-component of the far-field radiation from Shape D, which is associated with the charge distributions induced in the rectangle. When we draw arrows from the $+$ marks to the $-$ marks along the y-axis, we find that adjacent arrows are always directed oppositely, indicating that the y-component of the far-field radiation is externally small. In other words, Shape D behaves as a quadrupole regarding the y-component of the far-field radiation. It should also be noted that near-field components exist in the vicinity of the units in Shape D. With this fact in mind, we put the other metal nanostructure, Shape R, on top of Shape D. Through the optical near-fields in the vicinity of Shape D, surface charges are induced on Shape R. What should be noted here is that the arrows connecting the $+$ and $-$ marks along the y-axis are now aligned in the same direction, and so the y-component of the far-field radiation appears; that is, the stacked structure of Shape D and Shape R behaves as a dipole (Fig. 1.10c). Also, Shape D and Shape R need to be closely located to invoke such effects since the optical near-field interactions between Shape D and Shape R are critical. In other words, far-field radiation appears only when Shape D and Shape R are correctly stacked.

We fabricated structures consisting of (i) Shape D only, (ii) Shape R only, and (iii) Shape D and Shape R stacked. Although the stacked structure should ideally

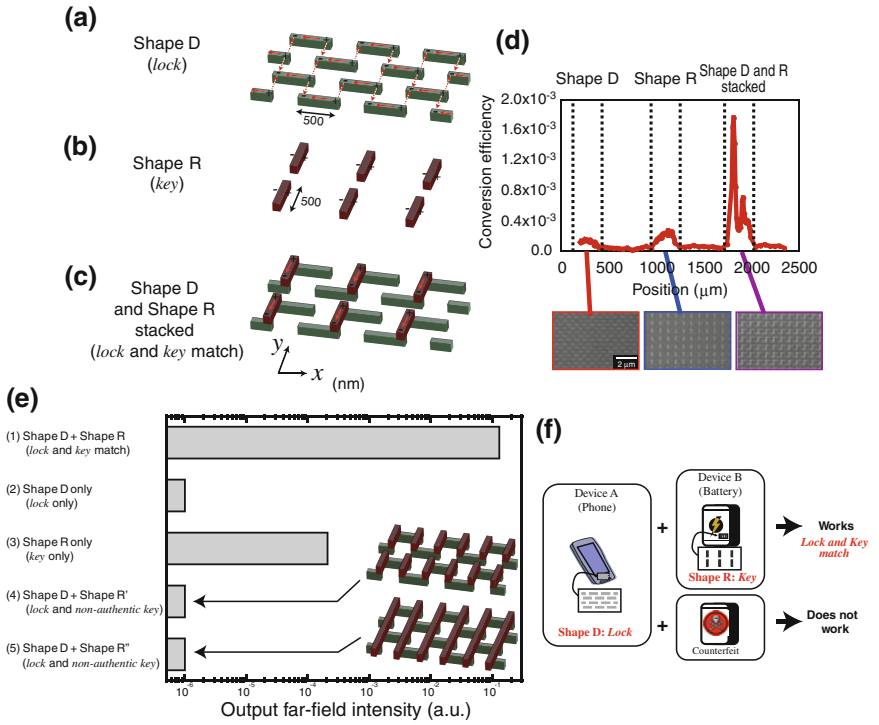


Fig. 1.10 Authentication, or *lock* and *key*, based on nano-optics. **a** Shape D, **b** Shape R, and **c** stacked structure of Shapes D and R. **d** Experimental demonstration of the lock and key matching, where a strong output signal appears only when Shape R matches Shape D. **e** Lower output signals when *non-authentic keys* are stacked on top of the *lock*. **f** An application of authentication based on nano-optics. Shape D on a mobile phone and Shape R on a certified battery

be formed by combining the individual single layer structures, in the experiment, the stacked structure was integrated in a single sample as a solid two-layer structure to avoid the experimental difficulty in precisely aligning the individual structures mechanically. The fabrication process was detailed in [21]. The lower side in Fig. 1.10d also shows SEM images of fabricated samples of (i), (ii), and (iii). The performance was evaluated in terms of the polarization conversion efficiency by radiating x-polarized light on each of the areas (i), (ii), and (iii) and measuring the intensity of the y-component in the transmitted light. The light source was a laser diode with an operating wavelength of 690 nm. Two sets of polarizers (extinction ratio 10^{-6}) were used to extract the x-component for the input light and to extract the y-component in the transmitted light. The intensity was measured by a lock-in controlled photodiode. Figure 1.10d shows the polarization conversion efficiency as a function of the position on the sample, where it exhibited a larger value specifically in the areas where the stacked structure of Shapes D and R was located, which agrees well with the theoretically predicted results.

From the theoretical standpoint presented in Sect. 1.3.1, what is important to obtain y-polarized output light is to locate an appropriately-designed Shape R on top of Shape D with high alignment accuracy so that coherently phased dipoles are induced in Shape R. As the theoretical results in Sect. 1.3.1 suggest, high-precision control and alignment is essential to induce the output light.

In Fig. 1.10e, we consider the output signals when we place differently-shaped structures on top of Shape D, instead of Shape R. That is, we intentionally put *non-authentic keys* on top of the lock. The far-field output light is evaluated by finite-difference time-domain (FDTD) electromagnetic simulations. The designs of Shape D and Shape R consist of arrays of gold rectangular units; each unit has a length of 500 nm, and a width and a height of 100 nm. As the material, we assumed the Drude model of gold with a refractive index of 0.16 and an extinction ratio of 3.8 at a wavelength of 688 nm. With Shape R′ and Shape R″, which are respectively represented in the insets of Fig. 1.10e, as well as with the *lock-only* and *key-only* shapes, the output signals do not appear, as shown from the fourth to the fifth row in Fig. 1.10e, since the condition necessary for far-field radiation is not satisfied with those shapes; namely, the correct *key* is necessary to unlock the *lock*.

Such functionality is useful, as mentioned in the beginning of this section, in ensuring hardware authentication, where a system should work only when the two nanostructures match. This would provide novel applications in the authentication of ubiquitous devices. For instance, the explosion of non-certified batteries is a serious issue in mobile devices [82]; the nano-optical lock-and-key principles and technologies described here could mitigate this problem by locating Shape D on the mobile phone and a corresponding Shape R on a certified battery, as schematically shown in Fig. 1.10f.

1.4 Stochastic Modeling of Near-Field Processes for Intelligent Material Formation

Precision control of the geometrical features of materials on the nanometer scale, such as their sizes and positions, are important factors in obtaining the intended functionalities of nanophotonic devices and systems in which multiple nanostructures interact via optical near-fields [11], and also for plasmonic devices [83]. For example, the sizes of QDs should be well-controlled to ensure that the quantized energy levels between adjacent QDs are resonant, to facilitate efficient optical near-field interactions [24]. Arrays of nanoparticles are important in various applications, such as nanophotonic devices [11], optical far-field to near-field converters, plasmonic light transmission lines [84], etc.

To satisfy such requirements, light-assisted, self-organized nanostructure fabrication principles and techniques have been developed [33, 34]. One example is the sol-gel synthesis of ZnO quantum dots (QDs) using photo-induced desorption, which yields QDs with reduced variations in diameter [33]. Another example of

light-assisted nanostructure fabrication is metal sputtering with light irradiation, which produces self-organized, size- and position-controlled metal nanoparticle chains [34]. In addition to the superior ability in regulating the geometries of nanostructures, these light-assisted, self-organized fabrication techniques have relatively simple experimental setups and have superior production throughput compared with, for instance, scanning-based methods, such as those based on scanning architectures [85, 86].

The physical mechanisms behind light-assisted nanostructure formation have been attributed to material desorption [33, 87, 88] or plasmon resonance between light and matter. However, stochastic physical processes are also present, as observed in the experimental data discussed below. Also, we consider that a stochastic approach is indispensable for taking account of the emergence of ordered structures and for the purpose of obtaining intended structures; in other words, for achieving "*nanointelligence*" in material production.

In light of this background, we approach light-assisted nanofabrication from a stochastic standpoint. Taking account of light-assisted processes, we build stochastic models that reproduce tendencies consistent with experimental observations. Through such considerations, we obtain critical insights into order formation on the nanometer scale, which will contribute to the design of nanophotonic devices and systems.

1.4.1 Light-Assisted Size-Regulation of Nanoparticles

First, we characterize the light-assisted, self-organized ZnO quantum dot formation, which was experimentally demonstrated in reference [33], with a stochastic approach. We first briefly review the experimental observations. In making ZnO QDs, synthetic methods using liquid solutions are advantageous because of their need for simple facilities and their high productivity [89] compared with other techniques [90, 91]. In conventional sol-gel methods [89], however, the size of the QDs varies by as much as 25 %. Liu et al. demonstrated a light-assisted sol-gel method that reduced the QD diameter variations [33]. When light with a photon energy higher than the bandgap energy is radiated during the ZnO QD formation process, electron–hole pairs could trigger an oxidation–reduction reaction in the QDs, causing the ZnO atoms depositing on the QD surface to be desorbed. In addition, such desorption is induced with a high probability when the formed QDs reach a particular diameter. This light-dependent QD size regulation has also been reported in other material systems, such as CdSe [87] and Si [88].

The insets in Fig. 1.11a, b respectively show transmission electron microscope (TEM) images of fabricated ZnO QDs without and with continuous-wave (CW) light illumination at a wavelength of 325 nm with a power density of 8 mW cm^{-3} [33]. Figures 1.11a, b respectively summarize the incidence rate of nanoparticles as a function of their diameter, whose variations decreased from 23 to 18 % with light irradiation. In particular, note that the diameter distributions are different between

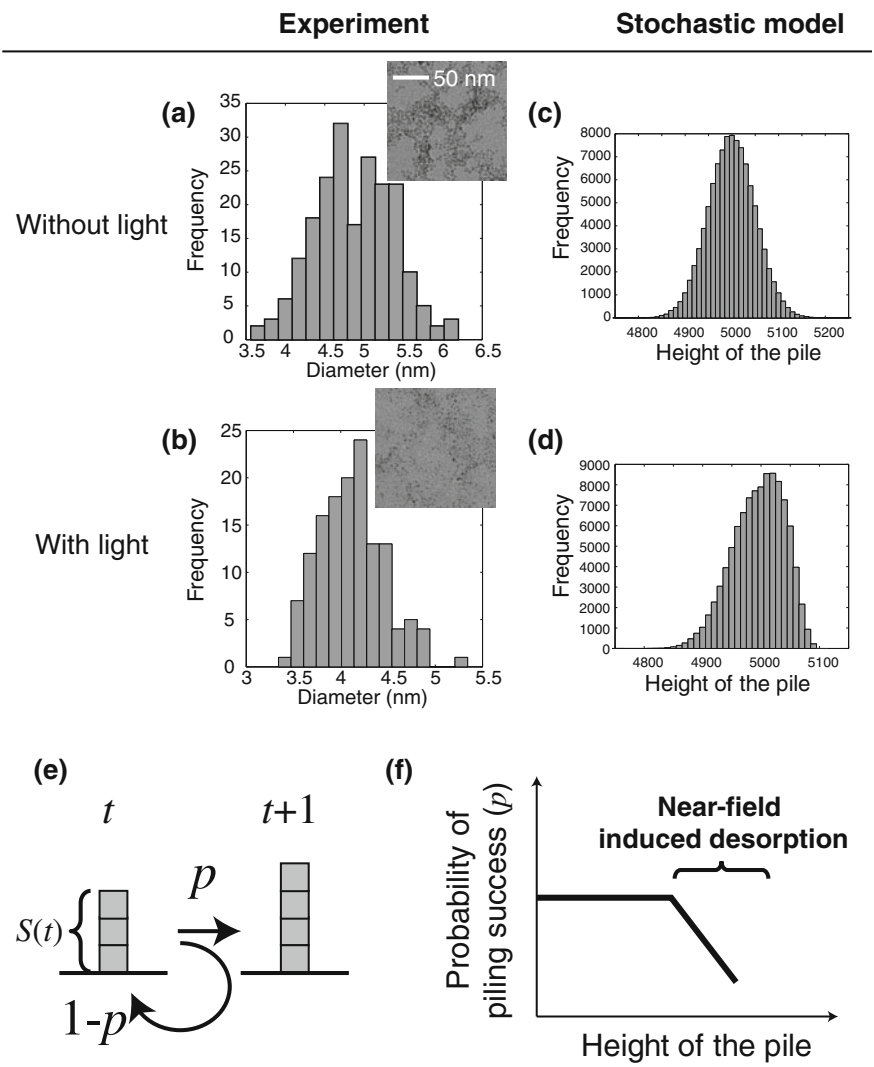


Fig. 1.11 Incidence patterns of the diameters of fabricated ZnO quantum dots (QDs) formed by a sol-gel method **a** without light irradiation and **b** with light irradiation. Insets in (**a**) and (**b**) are transmission electron microscope images of QDs without and with light irradiation, respectively. **c**, **d** Incidence patterns of the size distribution generated with the proposed stochastic model. The patterns are consistent with the experimental observations in (**a**) and (**b**). **e** A stochastic model of light-assisted nanoparticle formation. The growth of the QD is characterized with a one-dimensional pile-up model. The success of the piling depends on probability p. **f** The effect of light irradiation is modeled by a decrease in the probability p beyond a certain pile height, which corresponds to the diameter of the nanoparticle in the experiment

these two cases. Without light illumination (Fig. 1.11a), it exhibits behavior similar to a normal distribution. In contrast, with light irradiation, the distribution is skewed; in particular, the incidences at larger diameters decreased (Fig. 1.11b). We investigate this different behavior by means of stochastic modeling.

First, in the absence of light illumination, we represent the formation process with a statistical pile-up model, as schematically shown in Fig. 1.11e. An elemental material that constitutes a nanoparticle is represented by a square-shaped block. Such blocks are grown, or stacked one on another, with a piling success probability p; accordingly, the piling fails with a probability of $1 - p$. In other words, if we let the height of the pile at step t be $s(t)$, the piling probability is given by

$$\begin{cases} P\left[s(t+1) = s(t) + 1 | s(t)\right] = p \\ P\left[s(t+1) = s(t) | s(t)\right] = 1 - p. \end{cases} \tag{1.10}$$

Since this is equivalent to a random walk with drift, after repeating this process with an initial condition $s(0) = 0$, the resultant heights of the piles exhibit a normal distribution, as shown in Fig. 1.11c. Specifically, the statistics shown in Fig. 1.11c were obtained by repeating 10,000 steps for 100,000 different trials.

On the other hand, we model the effect of light irradiation in the formation process in the stochastic model as follows. As described above, since the material desorption is likely to be induced at a particular nanoparticle diameter, we consider that the piling success rate p is a function of the diameter, namely the height of the pile. For simplicity, we consider that p, which represents the deposition success probability, decreases linearly beyond a certain total pile height, as schematically shown in Fig. 1.11f. In other words, the material desorption is more likely to be induced beyond a certain pile size due to the resonant effect mentioned above. That is, the probability p in (1.10) is replaced with the following size-dependent probability:

$$p(s(t)) = \begin{cases} c & s(t) \le R \\ c - \alpha s(t) & s(t) > R \end{cases}, \tag{1.11}$$

where c and α are constants. With such a stochastic model, the resultant incidence distributions of the piles are skewed or reduced at larger sizes. In the calculated results shown in Fig. 1.11d, we assume $c = 1/2$ and $\alpha = 1/250$. The numerical results obtained through the statistical modeling are consistent with the experimental observations.

1.4.2 Light-Assisted Nanoparticle Array Formation

Yatsui et al. demonstrated self-organized formation of an ultralong array of nanoparticles based on near-field optical desorption (Fig. 1.12a, b) [34]. We first briefly describe the experimental observations. With conventional radio-frequency (RF) sputtering, aluminum was deposited on a glass substrate. A 100 nm-wide and

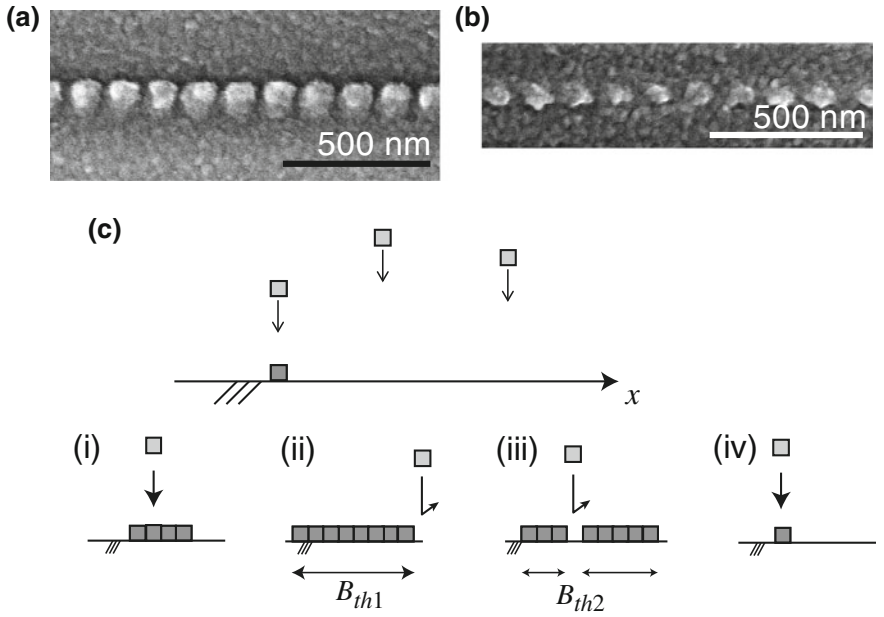


Fig. 1.12 An array of uniform-diameter, uniform-separation Al nanoparticles is self-organized along a groove, with **a** 2.33 eV light irradiation and **b** 2.62 eV light irradiation. **c** A stochastic model of the nanoparticle array formation. One-dimensional array in which an elemental block could be deposited at position x. (*i*)–(*iv*) Rules for successful deposition at a randomly chosen position x. (*i*) When the position belongs to one of the clusters, the cluster is maintained. (*ii*) Deposition is inhibited next to a cluster whose size is larger than B_{th1}. (*iii*) Deposition is inhibited at a position where the block sees clusters at both the left- and right-hand sides and when the total size of both clusters is larger than B_{th2}. (*iv*) In other cases, deposition at the position succeeds

30 nm-deep groove was formed in the substrate. Also, the substrate was illuminated with light linearly polarized perpendicularly to the direction of the groove during the RF sputtering. Thanks to the well-defined edge of the groove, a strong optical near-field was generated in its vicinity.

A metallic nanoparticle has strong optical absorption because of plasmon resonance that depends strongly on the particle size [92–94]. This can induce desorption of a deposited metallic material when it reaches the resonant diameter [95, 96]. It turns out that, as the deposition of the metallic material proceeds, the growth is governed by a tradeoff between deposition and desorption, which determines the particle diameter, depending on the photon energy of the incident light. Consequently, an array of metallic nanoparticles is aligned along the groove. While radiating continuous-wave (CW) light with a photon energy of 2.33 eV (wavelength: 532 nm) during the deposition of aluminum, 99.6 nm-diameter, 27.9 nm-separation nanoparticles were formed in a region as long as 100 μm (Fig. 1.12a).

As described above, the origin of the size regulation of the nanoparticles was attributed to the resonance between the nanoparticles and the illuminated light,

similarly to the case discussed in Sect. 1.4.1. At the same time, we consider that although this physical mechanism indeed plays a crucial role, it is not enough to explain the formation of the uniformly formed array structure. To explain such an observation, we need to extend the stochastic model described in Sect. 1.4.1 as follows.

In the modeling, we assume a one-dimensional horizontal system that mimics the groove on the substrate; it consists of an array of N pixels identified by an index i ranging from 1 to N. An elemental material to be deposited onto the system, experimentally by the RF sputtering described above, is schematically represented by a square-shaped block (Fig. 1.12c). The initial condition is a flat structure without any blocks.

At every iteration cycle, the position at which a block arrives, denoted by x, is randomly chosen. We determine the success of the deposition at x by the following rules. We denote the occupation by a block at position x of the groove by $S(x)$; $S(x) = 1$ when a block occupies a position x, and $S(x) = 0$ when there is no block at position x. Also, we use the term "cluster" to mean multiple blocks consecutively located along the groove. We also call a single, isolated block in the system a "cluster".

1. When the randomly chosen position x belongs to one of the cluster(s), namely, $S(x) = 1$, we maintain $S(x) = 1$ (Fig. 1.12c, i).
2. Even if $S(x) = 0$, when the chosen position x belongs to a "neighbor" of a cluster with a size greater than a particular number B_{th1}, the deposition is inhibited. That is, we maintain $S(x) = 0$ (Deposition is inhibited.) (Fig. 1.12c, ii).
3. Even if $S(x) = 0$, when the chosen position x has blocks at both its left and right sides and the total number of connected blocks is greater than B_{th2}, the deposition is inhibited. That is, we maintain $S(x) = 0$ (Fig. 1.12c, iii).
4. In other cases, the deposition at position x succeeds; namely, $S(x) = 1$ (Fig. 1.12c, iv).

The rules (ii) and (iii) correspond to the physical effect of the resonance between the material and the light illumination that facilitates desorption of the particle. The optical near-field intensity in the vicinity of a nanostructure follows a Yukawa function (1.6). Therefore, the optical near-fields promote material desorption, or in effect, inhibit material deposition, beyond a certain size of nanoparticles, which is characterized as rule (ii) above. Also, even when a single cluster size is small, meaning that the corresponding near-fields are small, when several such clusters are located in close proximity, a material desorption effect should be induced overall. Such an effect is represented as rule (iii) above. One remark here is that we do not pile more than two blocks at a single position x; that is to say, $S(x)$ takes binary values only, since our concern is how the clusters are formed in the 1D system.

Figure 1.13a shows the results of a numerical demonstration assuming a 1D array with $N = 1{,}000$. As statistical values in the simulations, we evaluated the incidence of the cluster size and the center-to-center interval between two neighboring clusters. Figure 1.13a summarizes the evolution of these two values at $t = 1{,}000$ and $t = 100{,}000$. In the numerical calculations, for the threshold values in rules (ii) and (iii), we assumed $B_{th1} = 8$ and $B_{th2} = 12$, respectively. We clearly observed that the

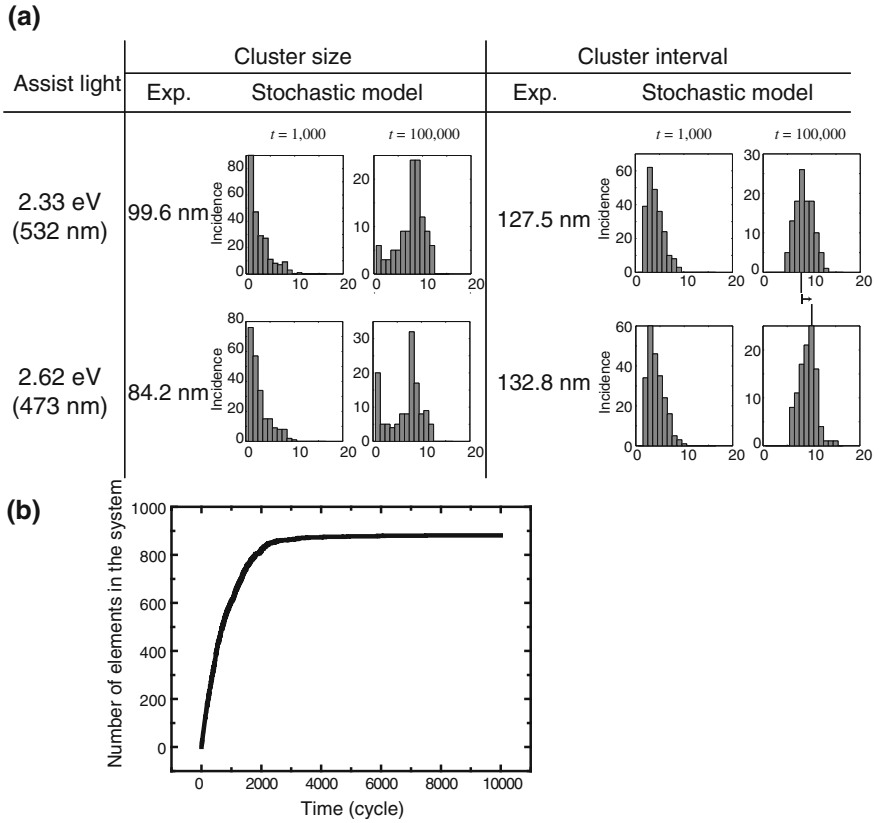


Fig. 1.13 **a** Evolution of the cluster size and the cluster interval based on a stochastic model. Both the size and the interval converge to incidence patterns that exhibit maxima at a particular value, which reproduced the experimentally observed size- and separation-controlled formation of a nanoparticle array. The interval of the nanoparticles is greater with higher photon energy. By modifying rule (ii) of the stochastic model, the cluster interval increases, which is consistent with the experimental observations. **b** Evolution of the number of elements in the stochastic model. After cycle $t = 3{,}000$, the number of elements is stable, meaning that the system is getting into a steady state in a self-organized manner

size and the interval converged to representative values, which is consistent with the experimental observations shown in Fig. 1.12a. Furthermore, we evaluate the total number of elemental blocks contained in the system as a function of elapsed time in the stochastic simulation. As shown in Fig. 1.13b it converges to a constant value beyond around 3000 iteration cycles, which is another indication that a self-organized process emerges based on optical near-field processes. Since the present modeling includes some parameters, this does not strictly exhibit so-called self-organized criticality [30]. However, we consider that the convergence to a uniformly

sized, uniformly separated 1D pattern is indeed a kind of self-organized critical phenomenon.

Moreover, as reported in reference [34], a similar experiment was conducted with a higher photon energy of 2.62 eV (473 nm) and an optical power of 100 mW, which resulted in the formation of 84.2 nm-diameter, 48.6 nm-separation nanoparticles (Fig. 1.12b). As summarized in the experimental results indicated in Fig. 1.13a, the diameter is slightly reduced and the nanoparticle distance is enlarged compared with the previous case of lower photon energy (2.33 eV (532 nm)). The reduced diameter of the nanoparticles is attributed to the fact that the higher photon energy leads to desorption at smaller diameters [33, 92]. The larger separation between adjacent nanoparticles is, however, not obviously explained.

We presume that a stronger light–matter resonance is induced at a higher photon energy, which more strongly induces material desorption, or inhibits the deposition of materials, in the neighboring clusters. We can take account of this effect by modifying the stochastic model described above. Instead of blocking the deposition at the neighboring positions by rule (ii), we consider that distant neighbors are also inhibited:

(ii′) Even if $S(x) = 0$, when x sees a cluster with a size greater than a particular number B_{th1}, within an area (a) between $x - 3$ and $x - 1$ or (b) between $x + 1$ and $x + 3$, the deposition is inhibited. That is, we maintain $S(x) = 0$.

While preserving B_{th1} and B_{th2} values with the previous example, the cluster size statistics evolve as shown in Fig. 1.13a. At the iteration cycles $t = 1{,}000$ and 100,000, the incidences of single-sized clusters are large. This is due to the strict inhibition rule (ii′) above, which reduces the chance of clusters growing. Treating such a single-sized cluster as an artifact, or a virtually ignorable element, in the system, we evaluate the cluster-to-cluster interval except for single-sized clusters. The cluster interval converges to a maximum of 10, as shown in Fig. 1.13a, which is larger than the previous case, which converged to 8. This is consistent with the experimental observations.

Such a stochastic modeling has also been applied in analyzing the morphologyof a solar cell that exhibits photosensitivity below the bandgap energy made by an optical near-field assisted method [31].

Furthermore, mathematical modeling helped to unveil the mechanisms of light-assisted surface flattening [35]. More specifically, a two-dimensional hierarchical surface roughness measure taking account of the hierarchical property of optical near-fields was developed and was found to be useful in a demonstration of light-assisted surface processing [99]. As such, system-oriented research will contribute to realizing “nanointelligence” in material applications.

1.5 Conclusion and Future Prospect

In this chapter, we have discussed nanophotonics from the standpoint of information, toward achieving intelligent functions, or what we call “nanointelligence”. In particular, we highlighted optical excitation transfer involving optical near-field

interactions keeping an eye on the possibility of going beyond von Neumann' computing. Nanophotonic security is also presented both theoretically and experimentally. Finally, stochastic modeling, as well as its contribution to material applications, was discussed.

Finally, we make a few remarks on the future prospects of research and development in nanophotonics from an information physics, or system-level viewpoint. We consider that future research and development will be roughly grouped into three categories. The first category concerns further investigation of basic and emergent properties in nanophotonics. For example, scalability and suitable information representations for nanophotonic systems should be further investigated in the future. The second category involves design-related concerns. This chapter mainly exploited optical energy transfer and multipole attributes offered by near-field interactions. The degrees-of-freedom available on the nanoscale, however, are in fact much more abundant and should be thoroughly investigated; these include magnetic field, electron spin, and so on. For instance, magnet-chiral effects [97], energy transfer in diluted magnet semiconductors [62], etc. suggest the importance of including a discussion of spin in the treatment of optical near-fields. In addition, morphology in nanoscale material enables interesting functions, such as light emission from indirect-type semiconductors (e.g., silicon) [15]. It has been suggested that coherent coupling between phonons and optical near-fields lies behind such phenomena [98]. Rigorous theoretical schemes, including ones that offer systematic design frameworks for realizing versatile materials and functionalities, should be investigated. The third category concerns applications; socially and industrially important applications will be pursued in the future [6, 7], and "intelligent" attributes will be one area where the work described in this chapter will contribute.

Finally, we hope that this chapter will contribute to stimulating research in the cross-disciplinary areas of nanophotonics, information, and materials, from fundamental principles to practical applications.

Acknowledgments The authors would like to thank many collaborators for illuminating discussions over several years, in particular H. Hori, T. Kawazoe, T. Yatsui, W. Nomura, K. Kobayashi, K. Akahane, N. Yamamoto, T. Inoue, F. Peper, K. Leibnitz, M. Hoga, Y. Ohyagi, T. Matsumoto, L. B. Kish, C. Lienau, E. Runge, S. Huant, and S.-J. Kim. This work was supported in part by Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (JSPS) and the Strategic Information and Communications R&D Promotion Programme (SCOPE) of the Ministry of Internal Affairs and Communications.

References

1. R.S. Tucker, R. Parthiban, J. Baliga, K. Hinton, R.W.A. Ayre, W.V. Sorin, J. Lightwave Technol. **27**, 243 (2009)
2. T. Ryhänen, M.A. Uusitalo, O. Ikkala, A. Kärkkäinen, *Nanotechnologies for Future Mobile Devices* (Cambridge University Press, Cambridge, 2010)
3. M. Naruse, N. Tate, M. Ohtsu, J. Optics **14**, 094002 (2012)
4. B. Javidi, *Optical and Digital Techniques for Information Security* (Springer, New York, 2005)

5. M. Ohtsu, K. Kobayashi, T. Kawazoe, S. Sangu, T. Yatsui, IEEE J. Sel. Top. Quantum Electron **8**, 839 (2002)
6. National Research Council, *Nanophotonics: Accessibility and Applicability* (National Academies Press, Washington, D.C., 2008)
7. MONA (Merging Optics and Nanotechnologies Association) consortium 2008 *A European roadmap for photonics and nanotechnologies*
8. M. Ohtsu, H. Hori, *Near-Field Nano-Optics* (Kluwer Academic/Plenum Publishers, New York, 1999)
9. D.W. Pohl, D. Courjon (eds.), *Near Field Optics* (Kluwer Academic, Dordrecht, 1993)
10. M. Ohtsu, K. Kobayashi, H. Ito, G.H. Lee, Proc. IEEE **88**, 1499 (2000)
11. M. Ohtsu, T. Kawazoe, T. Yatsui, M. Naruse, IEEE J. Sel. Top. Quantum Electron. **14**, 1404 (2008)
12. M. Ohtsu, K. Kobayashi, T. Kawazoe, T. Yatsui, M. Naruse, *Principles of Nanophotonics* (Taylor and Francis, Boca Raton, 2008)
13. T. Kawazoe, K. Kobayashi, M. Ohtsu, Appl. Phys. Lett. **86**, 103102 (2005)
14. H. Fujiwara, T. Kawazoe, M. Ohtsu, Appl. Phys. B **98**, 283 (2010)
15. T. Kawazoe, M.A. Mueed, M. Ohtsu, Appl. Phys. B **104**, 747 (2011)
16. T. Kawazoe, M. Ohtsu, K. Akahane, N. Yamamoto, Appl. Phys. B **107**, 659 (2012)
17. S. Yukutake, T. Kawazoe, T. Yatsui, W. Nomura, K. Kitamura, M. Ohtsu, Appl. Phys. B **99**, 415 (2010)
18. M. Aeschlimann, M. Bauer, D. Bayer, T. Brixner, F.A. Javier, W. Pfeiffer, M. Rohmer, C. Spindler, F. Steeb, Nature **446**, 301 (2007)
19. T. Kawazoe, M. Ohtsu, S. Aso, Y. Sawado, Y. Hosoda, K. Yoshizawa, K. Akahane, N. Yamamoto, M. Naruse, Appl. Phys. B **103**, 537 (2011)
20. T. Yatsui, S. Sangu, T. Kawazoe, M. Ohtsu, S.J. An, J. Yoo, G.-C. Yi, Appl. Phys. Lett. **90**, 223110 (2007)
21. N. Tate, H. Sugiyama, M. Naruse, W. Nomura, T. Yatsui, T. Kawazoe, M. Ohtsu, Opt. Express **17**, 11113 (2009)
22. N. Tate, M. Naruse, T. Yatsui, T. Kawazoe, M. Hoga, Y. Ohyagi, T. Fukuyama, M. Kitamura, M. Ohtsu, Opt. Express **18**, 7497 (2010)
23. A. Drezet, C. Genet, J.-Y. Laluete, T.W. Ebbesen, Opt. Express **16**, 12559 (2008)
24. M. Naruse, T. Kawazoe, R. Ohta, W. Nomura, M. Ohtsu, Phys. Rev. B **80**, 125325 (2009)
25. M. Naruse, T. Yatsui, H. Hori, K. Kitamura, M. Ohtsu, Opt. Express **15**, 11790 (2007)
26. C. Pistol, C. Dwyer, A.R. Lebeck, IEEE Micro **28**, 7 (2008)
27. C. Pistol, W. Chongchitmate, C. Dwyer, A.R. Lebeck, IEEE Micro **30**, 110 (2010)
28. P.B. Catrysse, S. Fan. Nano Lett. **10**, 2944 (2010)
29. N.G. Van Kampen (ed.), *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 2007)
30. P. Bak, C. Tang, K. Wiesenfeld, Phys. Rev. A **38**, 364 (1988)
31. M. Naruse, T. Kawazoe, T. Yatsui, N. Tate, M. Ohtsu, Appl. Phys. B **105**, 185 (2011)
32. M. Naruse, Y. Liu, W. Nomura, T. Yatsui, M. Aida, L.B. Kish, M. Ohtsu, Appl. Phys. Lett. **100**, 193106 (2012)
33. Y. Liu, T. Morishima, T. Yatsui, T. Kawazoe, M. Ohtsu, Nanotechnology **22**, 215605 (2011)
34. T. Yatsui, W. Nomura, M. Ohtsu, Nano Lett. **5**, 2548 (2005)
35. T. Yatsui, K. Hirata, W. Nomura, Y. Tabata, M. Ohtsu, Appl. Phys. B. **93**, 55 (2008)
36. T. Förster, Ann. Phys. **2**, 55 (1948)
37. S.A. Crooker, J.A. Hollingsworth, S. Tretiak, V.I. Klimov, Phys. Rev. Lett. **89**, 186802 (2002)
38. G.D. Scholes, G.R. Fleming, J. Phys. Chem. B **104**, 1854 (2000)
39. M. Kubo, Y. Mori, M. Otani, M. Murakami, Y. Ishibashi, M. Yasuda, K. Hosomizu, H. Miyasaka, H. Imahori, S. Nakashima, J. Phys. Chem. A **111**, 5136 (2007)
40. H. Tamura, J.-M. Mallet, M. Oheim, I. Burghardt, J. Phys. Chem. C **113**, 7548 (2009)
41. S. Sangu, K. Kobayashi, A. Shojiguchi, T. Kawazoe, M. Ohtsu, J. Appl. Phys. **93**, 2937 (2003)
42. T. Kawazoe, K. Kobayashi, J. Lim, Y. Narita, M. Ohtsu, Phys. Rev. Lett. **88**, 067404 (2002)
43. T. Itoh, M. Furumiya, T. Ikehara, C. Gourdon, Solid State Commun. **73**, 271 (1990)

44. H.J. Carmichael, *Statistical Methods in Quantum Optics 1* (Springer-Verlag, Berlin, 1999)
45. X. Wang, N. Wada, Opt. Express **15**, 7319 (2007)
46. J.W. Goodman, *Introduction To Fourier Optics* (Roberts & Company, Colorado, 2004)
47. I. Arsovski, T. Chandler, A. Sheikholeslami, IEEE J. Solid-State Circuits **38**, 155 (2003)
48. P.-F. Lin, J.B. Kuo, IEEE J. Solid-State Circuits **36**, 666 (2001)
49. M. Naruse, T. Miyazaki, F. Kubota, T. Kawazoe, K. Kobayashi, S. Sangu, M. Ohtsu, Opt. Lett. **30**, 201 (2005)
50. T. Kawazoe, K. Kobayashi, S. Sangu, M. Ohtsu, Appl. Phys. Lett. **82**, 2957 (2003)
51. K. Akahane, N. Yamamoto, M. Naruse, T. Kawazoe, T. Yatsui, M. Ohtsu, Energy transfer in multi-stacked InAs quantum dots. Jpn. J. Appl. Phys. **50**, 04DH05 (2011)
52. A. Shojiguchi, K. Kobayashi, S. Sangu, K. Kitahara, M. Ohtsu, Superradiance and dipole ordering of an N two-level system interacting with optical near fields. J. Phys. Soc. Jpn. **72**, 2984–3001 (2003)
53. M. Naruse, H. Hori, K. Kobayashi, T. Kawazoe, M. Ohtsu, Pulsation optical pulsation mechanism based on optical near-field interactions. Appl. Phys. B **102**, 717–723 (2011)
54. W. Nomura, T. Yatsui, T. Kawazoe, M. Ohtsu, J. Nanophoton. **1**, 011591 (2007)
55. T. Yatsui, H. Jeong, M. Ohtsu, Appl. Phys. B **93**, 199 (2008)
56. M. Naruse, H. Hori, K. Kobayashi, P. Holmstrom, L. Thylen, M. Ohtsu, Opt. Express **18**, A544 (2010)
57. W. Nomura, T. Yatsui, T. Kawazoe, M. Naruse, M. Ohtsu, Appl. Phys. B. **100**, 181 (2010)
58. M. Ohtsu, K. Kobayashi, *Optical Near Fields* (Springer, Berlin, 2004)
59. S. Haykin, *Communication Systems* (John Wiley & Sons, New York, 1983)
60. L.B. Kish, IEE Proc. Circuits Devices Syst. **151**, 190 (2004)
61. M. Naruse, T. Kawazoe, S. Sangu, K. Kobayashi, M. Ohtsu, Opt. Express **14**, 306 (2006)
62. K. Ohmori, K. Kodama, T. Muranaka, Y. Nabetani, T. Matsumoto, Phys. Status Solidi C **7**, 1642 (2010)
63. J. Seufert, G. Bacher, H.J. Schömig, A. Forchel, L. Hansen, G. Schmidt, K.W. Molenkamp, Spin injection into a single self-assembled quantum dot. Phys. Rev. B **69**, 035311 (2004)
64. M. Naruse, P. Holmström, T. Kawazoe, K. Akahane, N. Yamamoto, L. Thylén, M. Ohtsu, Appl. Phys. Lett. **100**, 241102 (2012)
65. M. Naruse, M. Aono, S.-J. Kim, T. Kawazoe, W. Nomura, H. Hori, M. Hara, M. Ohtsu, Phys. Rev. B **86**, 125407 (2012)
66. M. Aono, M. Hara, K. Aihara, Commun. ACM **50**, 69 (2007)
67. M. Aono, L. Zhu, M. Hara, Int. J. Unconventional Comput. **7**, 463 (2011)
68. B. Korte, J. Vygen, *Combinatorial Optimization: Theory and Algorithms* (Springer-Verlag, Berlin, 2012)
69. M. Aono, M. Naruse, S.-J. Kim, M. Wakabayashi, H. Hori, M. Ohtsu, M. Hara, Langmuir **29**, 7557 (2013)
70. S.-J. Kim, M. Naruse, M. Aono, M. Ohtsu, M. Hara, Sci. Rep. **4**, 2370 (2013)
71. R.L. van Renesse, *Optical Document Security* (Artech House, Boston, 2005)
72. M. Naruse, H. Hori, K. Kobayashi, M. Ohtsu, Opt. Lett. **32**, 1761 (2007)
73. N. Tate, W. Nomura, T. Yatsui, M. Naruse, M. Ohtsu, Opt. Express **16**, 607 (2008)
74. M. Naruse, T. Yatsui, W. Nomura, N. Hirose, M. Ohtsu, Opt. Express **13**, 9265 (2005)
75. M. Naruse, T. Yatsui, T. Kawazoe, N. Tate, H. Sugiyama, M. Ohtsu, Appl. Phys. Exp. **1**, 112101 (2008)
76. M. Naruse, T. Inoue, H. Hori, Jpn. J. Appl. Phys. **46**, 6095 (2007)
77. E. Wolf, M. Nieto-Vesperinas, J. Opt. Soc. Am. A **2**, 886 (1985)
78. T. Inoue, H. Hori, Quantum theory of radiation in optical near field based on quantization of evanescent electromagnetic waves using detector mode. in *Progress in Nano-Electro-Optics IV*, ed. by M. Ohtsu (Springer, Berlin, 2005), pp. 127–199
79. S.P. McGrew, Hologram counterfeiting: problems and solutions. Proc. SPIE **1210**, 66 (1990)
80. H. Matsumoto, T. Matsumoto, IPSJ J. **44**, 1991 (2003)
81. M. Naruse, T. Yatsui, H. Hori, M. Yasui, M. Ohtsu, J. Appl. Phys. **103**, 113525 (2008)
82. http://www.spotafakephone.com/spot-fake-batteries.cfm. Cited 5 Feb 2012

83. E. Ozbay, Science **311**, 189 (2006)
84. L.A. Sweatlock, S.A. Maier, H.A. Atwater, J.J. Penninkhof, A. Polman, Phys. Rev. B **71**, 235408 (2005)
85. T. Ishikawa, S. Kohmoto, K. Asakawa, Appl. Phys. Lett. **73**, 1712 (1998)
86. T. Yatsui, T. Kawazoe, M. Ueda, Y. Yamamoto, M. Kourogi, M. Ohtsu, Appl. Phys. Lett. **81**, 3651 (2002)
87. T. Torimoto, S. Murakami, M. Sakuraoka, K. Iwasaki, K. Okazaki, T. Shibayama, B. Ohtani, J. Phys. Chem. B **110**, 13314 (2006)
88. H. Koyama, N. Koshida, J. Appl. Phys. **74**, 6365 (1993)
89. E.A. Meulenkamp, J. Phys. Chem. B **102**, 5566 (1998)
90. S. Besner, A.V. Kabashin, F.M. Winnik, M. Meunier, Appl. Phys. A **93**, 955 (2008)
91. H.Z. Wu, D.J. Qiu, Y.J. Cai, X.L. Xu, N.B. Chen, J. Crystal Growth **245**, 50 (2002)
92. T. Yatsui, S. Takubo, J. Lim, W. Nomura, M. Kourogi, M. Ohtsu, Appl. Phys. Lett. **83**, 1716 (2003)
93. A. Wokaun, J.P. Gordon, P.F. Liao, Phys. Rev. Lett. **48**, 957 (1982)
94. G.T. Boyd, T. Rasing, J.R.R. Leite, Y.R. Shen, Phys. Rev. B **30**, 519 (1984)
95. J. Bosbach, D. Martin, F. Stietz, T. Wenzel, F. Träger, Appl. Phys. Lett. **74**, 2605–2607 (1999)
96. K.F. MacDonald, V.A. Fedotov, S. Pochon, K.J. Ross, G.C. Stevens, N.I. Zheludev, W.S. Brocklesby, V.I. Emel'yanov, Appl. Phys. Lett. **80**, 1643 (2002)
97. B.A. van Tiggelen, D. Lacoste, G.L.J.A. Rikken, Phys. B **279**, 13 (2000)
98. Y. Tanaka, K. Kobayashi, Phys. E **40**, 297 (2007)
99. M. Naruse, T. Yatsui, W. Nomura, T. Kawazoe, M. Aida, M. Ohtsu, Appl. Phys. Lett. **102**, 071603 (2013)

Chapter 2
DNA Self-Assembled Nanostructures for Resonance Energy Transfer Circuits

Chris Dwyer, Arjun Rallapalli, Mohammad Mottaghi and Siyang Wang

Abstract This chapter describes our work toward building a molecular-scale integrated circuit technology based on DNA self-assembly. Distinct from its purpose in biology, we co-opt DNA to fold nanoscale substrates onto which we pattern optically active molecules into networks, or circuits. Unlike conventional computing paradigms founded on the principles of electron currents in metals and semiconductors, we employ a quantum mechanical transport mechanism called resonance energy transfer (RET) to convey signals. However, circuits of any interest for computing have been difficult to demonstrate due to an enormous design space with many, many degrees of freedom. To overcome this challenge we have developed a methodology for the design of RET circuits implemented on DNA nanostructures. First, we describe the general principles of RET circuits and DNA self-assembly, our design methodology, and then we conclude with two working examples to highlight the potential of this new technology.

2.1 Introduction

Miniaturization of electronic components over the decades since the invention of the integrated circuit has driven the limits of what is physically possible with a lithographic process to amazing results. Despite the many ways that contemporary integrated circuits can be used at the macro-scale, the physical limits of device performance at nano- and molecular-scales preclude their use in most biological and chemical settings. Although it is possible to build hybrid systems, e.g., a biological sensor interfaced to a computer, embedding computational resources deeply into everyday materials, cells, or the environment remains a challenge.

C. Dwyer (✉) · A. Rallapalli · M. Mottaghi · S. Wang
Duke University, Durham, NC, US
e-mail: dwyer@duke.edu

M. Naruse (ed.), *Nanophotonic Information Physics*,
Nano-Optics and Nanophotonics, DOI: 10.1007/978-3-642-40224-1_2,

Within this context, the field of DNA nanotechnology has emerged which connects knowledge from molecular biology to the study and synthesis of technological, engineered constructs with nano- and molecular-scale features. More narrowly, DNA nanotechnology enables the manufacturing of nanoscale "substrates" by self-assembly with atomic precision for patterning molecules and a throughput (i.e., number of substrates per run) which rivals the most advanced silicon foundries.

However, the benefits of DNA self-assembly for computer *systems* has yet to be realized in large part due to the lack of functional blocks, or devices, which can be easily manufactured this way. To address this shortcoming a growing number of researchers have turned their efforts to understand computation from a molecular perspective. It is from this vantage that we consider the use of photo-active molecules, or chromophores, as basic elements in a molecular-scale circuit. Similar to the flow of electron and hole currents in a metallic-semiconductor circuit, networks of chromophores can control the flow of molecular excitons. Resonance energy transfer (RET) is the mechanism by which excitons hop between chromophores and due to their band structure, defined by their spectral properties, this is often in only one direction, similar to current flow in a PN diode. Thus, we are left to discover what networks of chromophores will lead to interesting and relevant circuit behaviors with which we might one day use to build integrated molecular systems for computing. Our work to date has revealed several important network design principles and their experimental verification has confirmed that it is possible to apply engineering principles to achieve some behaviors. Still other behaviors which are desirable for logic circuitry, e.g., gain and feedback, remain to be fully verified by experiment. The present work describes our design methodology and several of the tools we have developed to explore this new space. For the purposes of this chapter we will focus on the design of combinational logic, synthesizable by DNA self-assembly, and highlight the mathematical foundation on which our models are built.

The remainder of the chapter is organized as follows. In Sect. 2.2, we establish the foundations for an integrated molecular system. In Sect. 2.3, we describe the DNA self-assembly process. In Sect. 2.4, we provide the theoretical background of RET logic devices. In Sect. 2.5, we describe a design process that we developed to find the best chromophore network for a given functionality. We expand on metrics and design rules we use in our design process in Sect. 2.6 and the automation methods in Sect. 2.7. In Sect. 2.8, we provide two working examples to illustrate the proposed design process and highlight the potential of this new technology. We provide a discussion of our design process in Sect. 2.9 and conclude in Sect. 2.10.

2.2 Foundations for an Integrated Molecular System

In order to implement a given logic functionality, chromophores must be structured into a precise network, or circuit, to control the flow of molecular excitons. The dynamics of exciton transport are dictated by three physical phenomena: light absorption, light emission or fluorescence, and resonance energy transfer. Similar

to wavelength division multiplexing (WDM), different wavelengths can be used to represent the inputs to, and outputs from, the circuit. Certain chromophores can be designated input nodes since they generate excitons only when excited by a very specific wavelength of light. Fluorescence of certain other chromophores can represent the outputs of the circuit since different chromophores can also emit photons of a very specific wavelength. An important consequence of WDM for RET circuits is that despite being much smaller than the wavelength of incident light, the circuit inputs and outputs can be independently probed and monitored without any form of spatially controlled light. For example, RET circuits work the same at the end of a fiber optic cable as they do under a microscope or in a cuvette. Thus, the optical diffraction-limit for conventional optical devices does not limit the density of RET circuits.

RET is a near-field interaction between the excited state and absorption dipole moments between two molecules and has a transfer efficiency that is highly dependent on their separation (typically 2–10 nm). Thus, the behavior of a RET network is dependent on the types of chromophores in the network, their placement and relative separations. Given the requirement that chromophores be spaced within only a few nanometers of each other to form RET circuits, we focus our tool chain on using DNA grids as scaffolds since they have been shown to form RET circuits for various purposes [1]. With DNA nanotechnology individual chromophores can be attached to any base, or end of a DNA strand, which leads to countless possible attachments sites. Moreover, with over 300 different types of chromophores commercially available for each attachment site, there are more than 10^{30} feasible networks. To search this vast design space for the best chromophore networks that exhibit a given logic functionality, it is necessary to employ an automated design process. To that end, we have developed a tool chain that provides a structured approach to finding such RET circuits.

Unfortunately, the design methods available prior to this work for designing RET networks have been *ad hoc* and limited to very small collections (<5) chromophores. However, the number of domain-specific applications, e.g., the molecular ruler [2], and others in protein structure determination [3], that will greatly benefit from our structured design process presented here is large.

2.3 DNA Self-Assembly

The precise binding rules of DNA enable the creation of nanostructures with minimum pitch on the order of a few nanometers. These nanostructures can be used to place and interconnect nanoscale components with molecular-scale precision. Thus, DNA self-assembly is an enabling technology for new computing paradigms [1].

The challenge in creating DNA nanostructures is to specify the appropriate DNA sequences such that the desired structure (geometry) forms and is thermodynamically stable. To meet this challenge, DNA self-assembly can exploit the common technique of composing a small set of relatively simple pieces to create more sophisticated

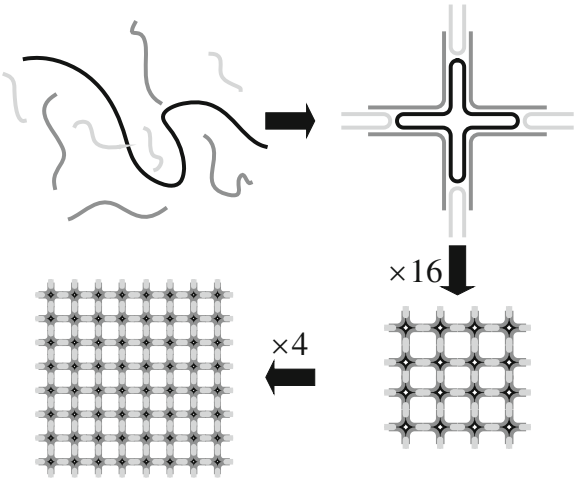


Fig. 2.1 The hierarchical approach to build DNA grids starts with nine single-strands of DNA (*top left*) which come together to form a cruciform motif or a *tile* (*top right*). Once individual tiles are formed, they can be assembled together to form *grids* (*bottom right*). Multiple grids can subsequently be assembled into larger grids (*bottom left*)

structures. The structure is composed through a *hierarchical* assembly of motifs. Ultimately, the final assembly step combines all motifs to form a grid structure.

For DNA there are many possible motifs, however we focus on only a few in the context of our energy transfer logic [4]. One class of motifs include junctions that enable three or more double stranded helices of DNA to interact and thus form specific structures (e.g., a triangle, a corner, etc.) Another important motif is a single strand of DNA protruding from a double stranded helix called a *sticky-end*. Two motifs with complementary sequences on their sticky-ends will bind to form a composite motif. These composite motifs may also have embedded sticky-end motifs and thus can also bind with other composite motifs to form another, larger, composite motif. This results in a hierarchical structure for motifs.

The DNA grid shown in Fig. 2.1 (bottom left) can be fabricated using hierarchical self-assembly of smaller structures called *tiles* (top right). Each tile consists of nine DNA single-strands whose sequences are designed such that they form a cruciform motif with sticky-ends on each extension of the cruciform. These sticky-ends help enable multiple tiles to assemble together to form DNA grids. The DNA grids act as a molecular breadboard enabling the attachment of molecules such as chromophores with full addressability.

The atomic precision, i.e., full addressability, of a molecular pattern on a DNA substrate is alone a remarkable achievement even when compared to contemporary high-end lithographic processing. The real advantage of the process, however, relies on the scale with which structures can be produced. Lithography can produce many, many copies from a mask set to drive per-unit costs down, as can DNA self-assembly by chemical manufacturing. Unlike high-end silicon processing which requires a multi-billion USD foundry, DNA self-assembly leverages the widespread availability of industrial chemical infrastructure requiring little non-recoverable expense. For

example, a 14 nm CMOS/FinFET foundry producing 80,000 wafers per month can be out-produced in a matter of hours by a single student using DNA self-assembly.[1]

2.4 RET Circuits and Logic Devices

RET logic devices are built from networks of closely spaced chromophores, or RET circuits, that direct the flow of excitons in a particular manner. The inputs and outputs for these devices are interfaced through absorption and emission of wavelengths of light, the presence and absence of a particular wavelength representing a true and false value, respectively.

2.4.1 Inputs and Outputs

When a photon is absorbed by a chromophore the photon's energy can promote a ground-state electron to an *excited* state, creating a molecular exciton (i.e., a tightly bound electron-hole pair). This process can only occur when the energy of the photon is equal to the singlet bandgap of the chromophore, which depends on the detailed molecular structure of the chromophore. For example, a chromophore called *Nile Red* has a bandgap that matches the energy of a 485 nm wavelength photon and will thus be maximally excited by blue light [5]. In a RET circuit, certain chromophores can be chosen such that their peak absorption wavelengths match the wavelengths of the incident input light sources. These chromophores can then be designated as the input nodes to each circuit.

An excited chromophore can subsequently *de-excite* through a variety of processes. One such de-excitation process is fluorescence, or the emission of a photon. The wavelength of emission is also dependent on the band structure of the chromophore. For example, *Nile Red* has a peak emission wavelength of 525 nm.[2] Fluorescence of chromophores provides a way of observing the output of a RET circuit. Instruments such as single-photon avalanche diodes and spectral filters enable the fluorescence of specific output chromophores in a RET circuit to be detected.

---

[1] This example uses projections for the TSMC Fab 15 (2015–2016) output of ≈80,000 450 mm wafers/month. The 14 nm process has been demonstrated on 6T SRAM cells, more generally NAND/INV cells, which occupy a footprint of 200 nm × 470 nm. That is, ignoring yield loss, Fab 15 could produce 1.35×10^{17} AND-equivalent circuits per month. DNA self-assembly synthesis executed by a student operating at 10 uM DNA grid concentration, each patterned with 24 AND gates, can assemble 1 mL of material. That is, ignoring yield loss, this student could produce 1.44×10^{17} AND-equivalent circuits per eight-hour run, or ≈20-fold higher per month.

[2] It is also important to note that a chromophore can absorb and emit photons in a range of wavelengths due to the existence of vibrational and rotational states for each electronic state.

2.4.2 Signal Transport

An excited chromophore can also de-excite via a non-radiative process called resonance energy transfer (RET). When an unexcited chromophore (the *acceptor*) is in close proximity (2 nm to 10 nm) to an excited chromophore (the *donor*) the exciton can quantum mechanically hop between the two. Similar to electron or hole diffusion across a junction, RET provides a mechanism for signal transport between the nodes of a RET circuit.

In a large population of donor-acceptor pairs, the fraction of the excited donor population that transfers energy to their unexcited acceptors defines the *transfer efficiency* of RET in that system.

$$TE = \frac{1}{1 + (\frac{r}{R_0})^6} \tag{2.1}$$

The transfer efficiency, as expressed in 2.1 and originally derived by Förster [6], depends on the sixth power of the separation (r) between the chromophores. The distance at which transfer efficiency is 50 % is defined as R_0, or the Förster radius for the pair of chromophores. The Förster radius depends on a variety of factors, the most important of which are the *spectral overlap* between the donor's emission and the acceptor's absorption profile and the relative *orientation* of the transition dipole moments of the two chromophores. Given commercially available chromophores and the sites available on a DNA nanostructure, virtually any given transfer efficiency is possible, in principle even very high efficiencies (e.g., >99.99 %).

2.4.3 Non-linear Transfer

RET is only possible from an *excited* donor to an *unexcited* acceptor. Specifically, at any given point in time an acceptor in an excited state cannot accept energy from any other donor, rendering the transient transfer efficiency between the pair zero percent.[3] This behavior can be used to induce changes in the transfer rate of a RET pair with the help of another donor. When an acceptor is shared among multiple donors it can be excited by only one of them at a time. This creates a competition among the donors and results in unsuccessful transfers from some of the donors to the shared acceptor. In this competition, the donor with higher transfer efficiency dominates the transfer to the acceptor; the acceptor becomes *saturated*, or occupied, from the perspective of the other donors.

Consider the case of three chromophores M, G, and N as depicted in Fig. 2.2. The left side of the figure shows the case where M is excited by photons; N and G

[3] In certain rare cases, RET can still occur if the energy difference between the first excited state and the second excited state of the acceptor chromophore matches the de-excitation energy of the donor.

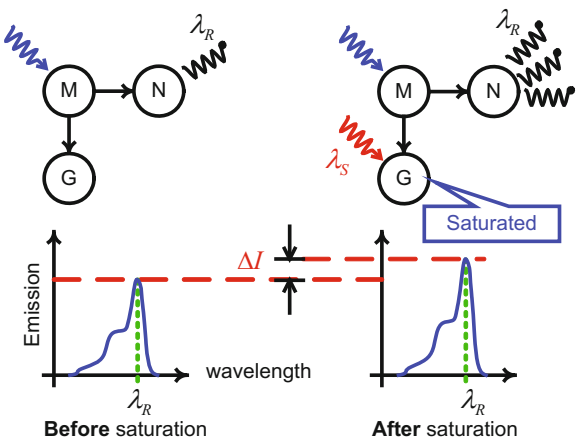


Fig. 2.2 *Left:* Node M is excited by photons and nodes N and G are excited by RET from M . *Right:* Nodes M and G are excited by photons. In this case, if the intensity of the light source for G is high enough, G appears *saturated* to M, thus no RET occurs from M to G. This results in higher effective transfer efficiency from M to N, with an increase of ΔI in I_R intensity

are excited via RET from M. In this case, none of the chromophores are saturated. The right side of the figure shows the case where both M and G are directly excited by photons. This creates a competition between two excitation processes for G, i.e., photon absorption and RET from M. With a high-intensity light source, G can be saturated such that it can never be excited by M. In this case, the effective transfer efficiency between M and G is zero. This results in higher effective transfer efficiency from M to N. The saturation effect enables a nonlinear control of molecular excitons in a RET circuit. This mechanism for nonlinearity offers a way to implement many kinds of devices, including the sequence detector discussed in Sect. 2.8.2.

2.4.4 RET-Based Logical Functions

RET-based signal transfer coupled with the non-linear transfer mechanism from acceptor saturation enables the design of RET circuits to implement any logic function [1]. Figure 2.3 shows the RET network of a two-input OR gate in which arrows show the direction of RET. As seen in the figure, node O is excited by either node

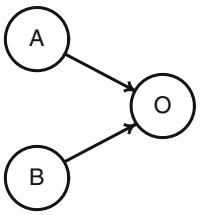


Fig. 2.3 A RET network that implements a two-input RET-logic OR gate. Node O fluoresces if either A or B is excited and transfers energy to O through RET

A or node B. In this network, logical 1 on A and B are represented by two distinct wavelengths that excite nodes A and B, respectively.

More generally the implementation of a logic function reduces to the design of a RET network that satisfies a set of requirements that are defined by the desired function. A structured and automated design process can facilitate the search for the RET network that best matches a set of user-defined design requirements.

2.5 Design Flow

The focus of this paper is to automate the component level design flow rather than the larger, system level flow. We consider components such as fundamental logic gates (OR and AND gates) as well as small- to medium-scale combinational logic. These components can be cascaded at a higher design-level to create more sophisticated logic circuits that can in turn be used in the design of arbitrary digital circuits. At the component level, the ultimate goal is to find a RET network that implements some desired logic function. Finding a RET network involves selecting the right chromophore-types for the nodes and tuning the distances between them.

The proposed design flow consists of six steps, as illustrated in Fig. 2.4 . The first two steps are defined by the user and are specific to the target application. Initially, based on the operating principles of RET networks, the user identifies the form of the target RET network that, under ideal conditions, expresses the desired logic function. In an ideal network, only the inter-nodal RET directions are considered. For example, the OR gate shown in Fig. 2.3 describes a network comprised of three nodes, two of which (A and B) are designated as the input nodes and are able to achieve RET with the third node (O). Ideally, each of the input nodes is exclusively excitable using two different wavelengths of light and the output node O is exclusively detectable using photo detection techniques.

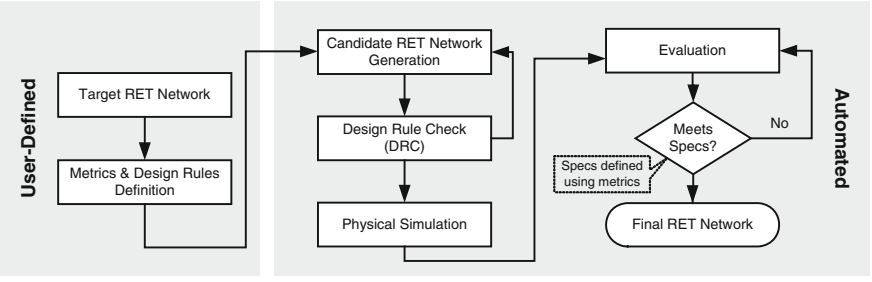


Fig. 2.4 Flowchart of the proposed design process: The first step is to identify the target RET network. The next step is to define the design rules and metrics which are application-specific. The next two steps prepare the set of candidate networks by exploring the sample space and determining those that pass the design rules check. Afterwards, the candidate networks are simulated and evaluated based on a set of user-defined metrics to find the final RET network

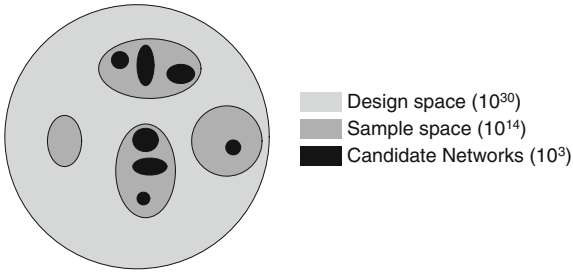


Fig. 2.5 Preparation of the set of candidate networks: Only parts of the huge design space are sampled. Each sample that passes the DRC qualifies as a candidate

The user then establishes a set of design rules that define the interactions between the nodes of the target RET network. Each design rule is an arbitrary function of a set of attributes that can be evaluated for an individual chromophore, a pair of chromophores, or the network as a whole without the physical simulation of the network. Finally, the user defines the metrics that are used in the development of specifications, which in turn are used to evaluate RET networks against the design requirements in a later step. The framework for defining the metrics and design rules is further explained in Sect. 2.6. With this, the behavioral representation of the desired application is established; the remaining steps in the design flow are fully automated.

The third and fourth steps explore the design space and identify candidate RET networks. As illustrated in Fig. 2.5, the prohibitively large size of the design space makes it impossible to perform a full exploration of the entire design space. Therefore, domain-specific heuristics have to be applied to guide the exploration (step 3) to avoid poor regions of the design space and only sample regions that are likely to contain a candidate network (sample space). All the networks in the sample space are checked against the design rules from step 2 (DRC) to form a set of candidate networks.

The next step is the physical simulation to find the realistic behavior of candidate networks followed by an evaluation step to determine which candidates meet the specifications. The design flow culminates with the identification of the RET network(s) that best matches the behavior of the target network identified in the first step of the design flow.

2.5.1 Comparison with Other EDA Design Flows

In the design automation of conventional CMOS logic devices, design rule checking (DRC) is applied to a possible sample layout out of many possible functionally-equivalent layouts of a single design. This step ensures the manufacturability of the device based on a target technology node. Similarly, the DRC step in the design of RET networks is applied to a sample network out of many possible networks in the design space.

2.5.2 RET Network Design as a Type-and-Site Assignment Problem

The RET network identified at the end of the design flow is eventually realized as a physical network of chromophores assembled on a nanostructure which maintains approximately-fixed positions. The underlying nanostructure can be a DNA grid or other similar chemical structure. These underlying structures impose limitations on the placement of molecules, disallowing certain positions. For example, on a DNA nanostructure, chromophores can only be attached to the bases and ends of the DNA strands implying the set of possible attachment sites is dictated by the host strand. This means that a network node cannot take on arbitrary (and uncorrelated) x-, y-, and z-values but rather it can be assigned one of the several possible sites, which in turn determines the specific coordinates of that node.

Other restrictions on physical RET networks stem from non-ideal spectral properties of available chromophores. Specifically, there exist only a limited number ($\approx$300) of chromophore types that can be placed on the underlying nanostructure. Each chromophore type (also called a dye) has unique excitation and emission spectra dictated by its molecular structure. For example *Lucifer Yellow* is a dye that has an emission peak at the 535 nm wavelength and an excitation peak at 428 nm [5]. These two key wavelengths (428 nm and 535 nm) cannot be customized and are dictated by the molecular structure of *Lucifer Yellow*. Thus, it is not the case that every desired spectral property can be found in commercially available dyes. Furthermore, the unique spectral properties of a dye may increase the complexity of the design process by introducing new considerations; e.g., the parasitic excitation of a chromophore due to (an unwanted but present) spectral overlap with another chromophore.

Therefore, RET network design can be viewed as finding the optimal "*type-and-site assignment*" for every node of the ideal RET network. The optimal assignment is important to minimize the deviation of the physical RET network from the behavioral ideal. Depending on the desired logic function, it might also be necessary to enhance the network functionality by finding the optimal excitation wavelengths for each node such that parasitic excitations are minimized.

2.6 Metrics and Design Rules

The RET-logic design flow presented here involves the validation of many candidate RET networks. A candidate RET network is the network that corresponds to a possible type-and-site assignment for each node. To validate a candidate RET network, its behavior is simulated and compared with that of the ideal RET network. Simulation of a RET network is typically very slow and the number of possible sample networks is very large (e.g., 10^{14}). Therefore identifying *disqualified* sample networks without simulating them can noticeably expedite the design process. Design rules are meant for this purpose. In addition to design rules, some metrics are also required to enable

ranking of *qualified* candidate networks after simulation. Metrics enable meaningful comparison between candidate RET networks. They are typically extracted from simulation results and their specific definition depends on the desired logic function, e.g., the propagation delay of a signal defined as a metric.

2.6.1 Metrics

Metrics are behavioral attributes of the network that can be used to rank simulated network behavior. The metrics are extracted from simulation results and are user defined. The simulation provides time-resolved details about the evolution of the excited population at each node as well as its fluorescence, RET, etc., which can be used in the extraction of a metric. For example, the evolution of fluorescence from a node can be processed to find the decay rate of that node (similar to an RC decay constant). Furthermore, metrics are also used in the definition of specifications which in turn are used in the evaluation of simulated RET networks.

2.6.2 Design Rules

Design rules are sets of rules that can disqualify a candidate network without simulation. Disqualifying type-and-site assignments can be identified by spectral and spatial design rules, respectively. For example, a minimum or maximum required distance between two nodes in the network can be declared as a spatial design rule. While the set of design rules highly depends on the target logic function and its specific ideal RET network, each individual rule is an arbitrary function of a set of attributes that can be evaluated for an individual dye or site or a pair of them.

The following is a list of attributes that we have found are useful in RET design rules.

Intrinsic Lifetime In a solution of same-type chromophores, when a population of molecules is excited they do not de-excite simultaneously. In fact, the de-excitation of the population resembles the discharge of a capacitor in an RC circuit for which a time-constant is defined. The time-constant for the de-excitation of a population of excited chromophores (in the absence of any acceptors) is called the intrinsic lifetime. For example an intrinsic lifetime of 5 ns means that 63 % of the population of chromophores de-excites 5 ns after being excited. Typical lifetime values are listed in Table 2.1 denoted by τ.

Extinction Coefficient Extinction coefficients are a measure of how strongly a dye absorbs light at a particular wavelength. It can be viewed as the efficiency of absorption at a given wavelength.

Quantum Yield An excited chromophore can be de-excited without emitting a photon, i.e., the energy dissipates non-radiatively. The quantum yield of a dye is

Table 2.1 Typical quantum yield and lifetime values

Chromophore name	QY	τ (ns)
Alexa Fluor 488	0.92	4.1
Alexa Fluor 532	0.61	2.5
Alexa Fluor 546	0.79	4.1
Alexa Fluor 555	0.10	0.3
Alexa Fluor 568	0.69	3.6
Alexa Fluor 594	0.66	3.9
Alexa Fluor 647	0.33	1.0
Alexa Fluor 660	0.37	1.2
Alexa Fluor 680	0.36	1.2
Alexa Fluor 700	0.25	1.0
Alexa Fluor 750	0.12	0.7

defined as the fraction of absorbed photons that are emitted as a result of de-excitation. Typical quantum yield values are listed in Table 2.1 denoted by QY.

Spectral Separation The separation between the excitation peaks of a pair dyes defines their spectral separation. The excitation peak is the wavelength at which a given dye exhibits the strongest light absorption.

Förster Radius (R_0) It is the donor-to-acceptor distance at which the FRET efficiency is 50 %. As discussed, it depends on the molecular structure of the donor and acceptor and each dye generally has a different R0 to every other dye.

Transfer Efficiency Not all donor excitation events result in the excitation of the acceptor through RET; instead, an excited donor may de-excite through another relaxation pathway, e.g., by emitting a photon. Transfer efficiency is the fraction of the excitation events that result in RET events and is calculated using 2.1.

2.7 Design Automation Process

Given the importance of RET and its applications in other fields, especially in fluorescence microscopy, there are a variety of tools available for forming the best RET pairs [7–9]. However, these tools are not sufficient for optimizing a RET network because they are *ad hoc* and use simplified interaction models (e.g., strong absorption at a particular wavelength) and no hierarchies which limits their use to only very small networks (i.e., maximum of two or three nodes). Even for small networks these tools cannot guarantee the best behavioral match between the desired logic function and the resulting network.

An alternative, but trivial, method is to randomly select candidate RET networks and evaluate them by simulation. In this method the computational effort for selection is low but there are obviously no guarantees on the optimality of the resulting design.

We have developed a set of tools to automate the design of arbitrary RET networks. These tools, each of which brings computer assistance to different levels of the design flow, can be chained either via direct access to the memory space or through well-defined text files.

2.7.1 Scripting Language for Tool-Chain Control

Design automation often requires run-time evaluation of application-specific mathematical expressions. Furthermore, on-the-fly execution of small user-defined scripts is required to automate the cascading of tools and flow from one tool to another. In order to obviate the need for code-compilation in such cases, we developed a scripting language to enable run-time execution of small user-defined programs called *proglets*, which have direct access to the memory space of each tool without any OS intervention for memory sharing. The *proglet* interpreter itself is compiled as part of each tool and therefore the memory space is naturally shared between the tool and the interpreter making memory access faster by avoiding slow system calls.

Each tool provides a set of library functions as well as uniquely-named data structures that users can access in their *proglets* to perform any arbitrary processing on the generated results; e.g., extracting the propagation delay of a signal after simulation.

2.7.2 Spectral Analysis Tool

The spectral analysis tool takes as input the set of spectral design rules and generates the set of qualified type-to-node assignments (i.e., half of the type-to-site problem with respect to user-defined template nodes). The design rules are expressed in the form of XML-formatted proglets. The following example qualifies donor-acceptor RET pairs (DA) with emission peaks after 450 nm and 660 nm for the donor and the acceptor, respectively.

```
<Dyes>
  <Set name="D" Filter="$D.EmPeak > 450" /> #450nmblue
  <Set name="A" Filter="$A.EmPeak > 660" />
</Dyes>
<Combine>
  <Pair join="$D,$A" name="DA"
     Filter="$D.ExcPeak<$A.ExcPeak && JOverlap ($D,$A)>1">
</Combine>
```

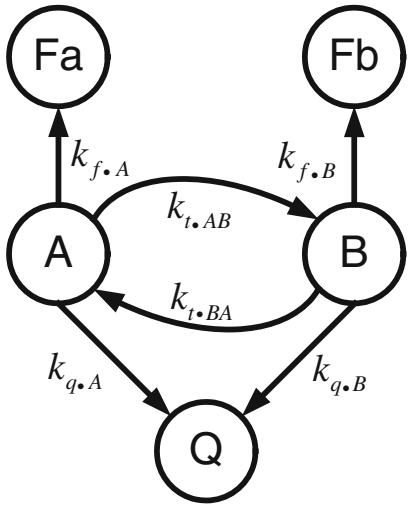
2.7.3 Geometry Analysis Tool

The geometry analysis tool takes as input the set of geometric design rules and generates the set of qualified site-to-node assignments (i.e., the other half of the type-to-site problem). The geometric design rules can be based on the geometry of an underlying nanostructure or a heuristic sampling of possible geometries. For example, the target RET network can be placed on a uniform grid in which the sampling of the site-to-node assignments is achieved by sampling different grid resolutions. For example, the following rules specify the spacing between nodes n1 and n3 are from 1 nm to 4 nm in steps of 0.1 nm.

```
Dist(n1, n3) := [1, 4, 0.1]nm
Dist(n3, n2) := [2, 6, 0.2]nm
```

2.7.4 Linear Approximation Tool

In general the behavior of RET networks with saturated or partially saturated acceptors is non-linear with respect to changes in the network and input conditions. However, the behavior of semi-linear RET networks can be approximated by a linear network which can be statistically described by a set of ordinary differential equations (ODE's). In such cases, standard ODE solutions can be used to obtain the network response with sufficiently low error. These solutions typically run noticeably faster than detailed, non-linear simulation methods and are therefore good for quickly pruning the search space. We use a continuous-time Markov chain (CTMC) solution to *approximate* the response of semi-linear networks. CTMC is a valid model to apply because exciton transfer in a network of chromophores are dictated by stochastic processes that are *memoryless*, i.e. the next state of an exciton in the network depends only on the present state, not on the sequence of state transitions that preceded it. In the CTMC model, each chromophore in the network is represented by two states in a Markov chain: transient state and absorbing state. Given an initial condition (denoted by the starting position of an exciton), a CTMC solution will result in time-resolved probabilities of any of the other states being occupied by the exciton. These probabilities can then be used to determine time-resolved relative fluorescence intensities of each chromophore. Since the output of a RET-based logic device is measured using fluorescence intensities, the CTMC model provides an analytical way of determining time-resolved output of a given network of chromophores. The linear-simulation results from CTMC are checked against a set of rules that are based on user-defined metrics to rapidly identify disqualified candidate networks. Figure 2.6 shows a Markov chain representation of a donor-acceptor pair of chromophores. The rates of transfer annotated on the figure are described by Förster's theory [6].



$$k_{f\bullet i} = Q_D \left(\frac{1}{\tau_D^o} \right)$$

$$k_{q\bullet i} = (1 - Q_D) \left(\frac{1}{\tau_D^o} \right)$$

$$k_{t\bullet ij} = \frac{3}{2} \frac{\kappa_{ij}^2}{\tau_D^o} \left(\frac{R_{0_{ij}}}{r_{ij}} \right)^6$$

Fig. 2.6 *A* and *B* are two chromophores in the system. *Fa* and *Fb* are the fluorescence absorbing states for *A* and *B*, respectively. *Q* is the universal absorbing state for quenching. κ_{ij}^2 is a measure of the orientation between the donor (*i*) emission dipole moment and the acceptor (*j*) absorption dipole moment. R_0 is the distance at which the RET transfer efficiency is 50 %; it depends on the spectral overlap between the donor emission and acceptor absorption. Q_D and τ_D are intrinsic quantum yield and fluorescence lifetime of the donor in the absence of acceptors respectively

2.7.5 Physical Simulation

Physical simulation of a RET network predicts the response of the manufactured network with a high degree of accuracy, similar to, e.g., PISCES for semiconductor device physics. In this section, we first describe related work in physical simulation of chromophore networks based on quantum mechanics and molecular dynamics. The development of these models was driven by applications such as RET-based molecular rulers and emulation of light harvesting complexes. Another family of chromophore network simulation discussed in this section is based on stochastic models such as Markov chain-based models (already described) and Monte Carlo models. Finally, the section will end with a description of a hybrid Monte Carlo/Markov chain model we developed, called SCIMM.

Related Work in Quantum Mechanical Models

Zhu et al. in [10], describe a simulation model of excitation energy transfer in supramolecular Pheophorbide-*a* chromophore complexes in solution. The detailed explanation of the expressions that are derived in this model is beyond the scope of this report but can be found in the corresponding reference. In this model, the Hamiltonian (energy operator) for the entire chromophore complex is defined as the sum of

the kinetic energy and the potential energy surfaces (PES) of all the chromophores in the complex. The PES of each chromophore characterizes the excitation of a given chromophore and includes the electrostatic coupling with all the other chromophores in the complex. Using this Hamiltonian, they solve for the eigenvalues or energy states that satisfy the Schrödinger's equation. These energy values, along with spatial information of the chromophore complex, are used to translate the molecular density to an electronic charge density, which was subsequently used to derive an expression for inter-chromophore coupling. Furthermore, they add in various molecular-level dynamics, such as vibrational motion of the chromophores, to the inter-chromophore coupling relation. They use this expression along with rate equations described by Förster's theory to come up with a time-dependent exciton density matrix. From this, time-resolved and frequency-resolved fluorescence information of the chromophores was derived.

Simulation of a given chromophore complex using this process entails running molecular dynamics simulations to obtain the vibrational and molecular mechanics components. Because the complexity of molecular dynamics increases with higher temperature, the quantum mechanical model described in [10] is only applicable if characteristic energy differences to be overcome in the excitation energy motion are less than or equal to the thermal energy defined by Boltzmann's relation. While this model can provide accurate time-resolved and frequency-resolved fluorescence information, it involves molecular dynamics simulations that are compute-intensive, limiting the scalability of the model to larger chromophore networks.

There are several other publications [11–14] that describe models that are based on quantum mechanics and molecular dynamics. Most of these publications state that the motivation for such models is to design artificial light harvesting complexes similar to those found in plants. For example, in [14], the authors derive a second-order time-convolution master equation from the principles of quantum mechanics. They conclude that this equation can be used to accurately estimate the energy transfer efficiency of light harvesting complexes. For the purpose of our application of finding the correct RET networks from a large design space, models based on quantum mechanics with molecular dynamics are too slow. Therefore, stochastic models that rely on the inherent probability distributions of exciton flow in RET networks are better suited.

Related Work in Stochastic Models

While the quantum mechanics and molecular mechanics approaches involve modeling of electronic wave functions in a mechanical force field, the stochastic models described in the following sections rely on the random nature of excitation of electrons in molecules. As described in the background section, the rates of fluorescence, quenching, and transfer to nearby molecules can be determined using Förster's theory and known or easily obtainable information about individual chromophores. These rates can be used in a random-walk process (Monte Carlo approach)

or in time-resolved energy migration dynamics solved using differential equations (Markov chain approach as described in Sect. 2.7.4).

Demidov in [15] describes a Monte Carlo based simulation tool to simulate energy migration in molecular complexes. In this mode, the user first defines the number of excitations to play through the simulation. Due to the law of large numbers, the accuracy of the model increases as more excitations are simulated. Based on the user-defined chromophore network settings, the model generates the spatial structure of the network of chromophores, including the coordinates of the chromophores and the coordinates of their absorption and transition dipole moments. The next step in the model is to calculate the pairwise transition probabilities using Förster's theory. The model then determines the number of absorbed photons using an incident flux setting and the absorption cross section of all the chromophores in the network. Then the next state for each excitation is randomly chosen based on the fluorescence, quenching, singlet-singlet annihilation, and transition probabilities. This process is repeated until all the excitons are resolved.

This approach allows for multiple excitation and de-excitation events to be processed simultaneously, thereby providing the capability to capture inter-chromophore effects. For example, in a three-chromophore system, an excited chromophore typically has four de-excitation paths, i.e., transfer to either chromophore, fluoresce, or quench. However, if one of the accepting chromophores also happens to be in an excited state, the former chromophore would only have three pathways, due to one of the accepting chromophores being saturated. Furthermore, the model takes into account the incident photon flux when determining the probability of singlet-singlet annihilation, another process which causes the behavior of a RET network to be non-linear.

It should be emphasized that the probabilities used in the Monte Carlo random walk are steady-state probabilities calculated from Förster's theory, i.e., the probability for a given de-excitation path (next state) is calculated by taking the ratio of the de-excitation path rate to the sum of the rates of all de-excitation paths. In reality, the probability for a path approaches a steady-state probability in a finite amount of time. As the time-step is decreased (to increase the accuracy of a random walk), the difference between the steady-state probabilities and the time-resolved probabilities (such as those determined from a Markov chain solution) increases.

Simulation of Chromophore Interactions Using Monte Carlo Methods and Markov Chains (SCIMM)

To combine the time-resolved probabilities from a CTMC solution with the ability to model non-linear effects of a Monte Carlo random walk, we have developed a hybrid model called SCIMM (pronounced "skim"). Figure 2.7 shows a functional block diagram of the Simulation of Chromophore Interactions using Monte Carlo methods and Markov chains (SCIMM) model. In SCIMM, the outer loop is a Monte Carlo loop that iterates through discretized time and keeps track of light absorption and fluorescence events. If for any given time step there exists an absorption event

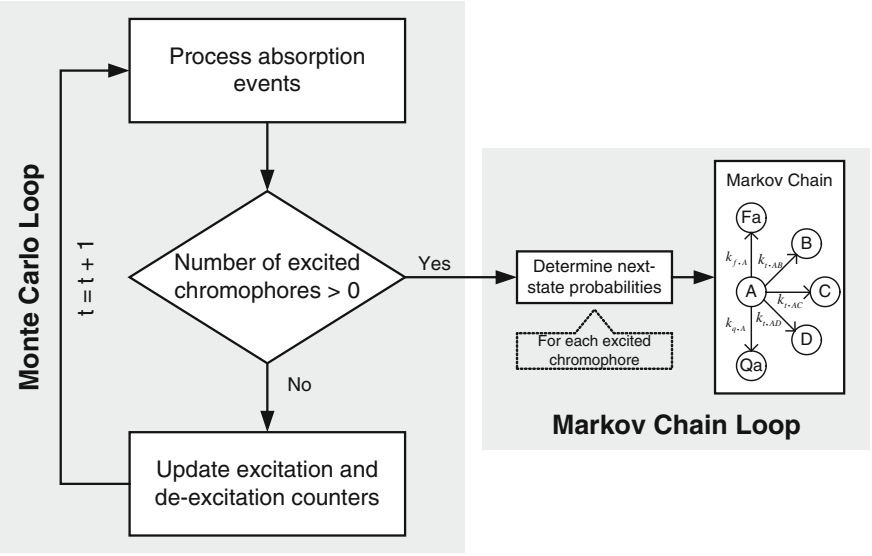


Fig. 2.7 Functional block diagram of the MCMC model. The SCIMM model consists of two loops, the outer Monte Carlo loop and the inner Markov Chain loop. For each time step, all the absorption events are first processed. If any excited chromophore are present in the system at the given time step, a CTMC model is setup for each excited chromophore to determine the next-state probabilities. These probabilities are used in a weighted random walk in the Monte Carlo loop. The excitation and de-excitation events of each chromophore in the system are recorded. Once the next states of all the excited chromophores are processed, the time step is incremented

or excited chromophore(s), we descend into an inner loop. The inner Markov Chain loop generates a CTMC model for each excited chromophore and determines the time-resolved probabilities of the next step for its corresponding exciton in a manner similar to sampling kernels used by Gibbs sampler. Once the next step for every excited chromophore is determined, we ascend into the Monte Carlo loop, update event counters, and increment to the next time step. This approach allows for the time-resolved probability distributions of next states to be captured within a discretized event-driven Monte Carlo simulation. As a result, the interaction between simultaneously excited chromophores and multiple excitons are captured as a histogram of events with respect to time. Therefore, the results produced by this simulation can be compared directly with those obtained from time-correlated single photon counter (TCSPC) measurements collected in the lab.

2.7.6 Batch Evaluation Tools

After the search space is sufficiently pruned, a set of candidate networks (all of which have passed DRC) is prepared for detailed physical simulation and evaluation. The

batch evaluation tool provides a platform for running simulations and performing on-the-fly evaluations which direct the flow of candidate network evaluation. This tool takes as input the list of all candidate networks (from search-space pruning) along with a list of parameters (e.g., excitation wavelengths) and invokes the appropriate simulator for each network. After the simulation it performs post-simulation analysis to extract the user-defined metrics. With this tool it is possible to skip the simulation of some candidates based on the metric values of previous simulations; for example, in the following snippet after the first simulation, metric @delay is extracted and if it is less than 10 the second simulation is skipped. In this example *$a*, *$b*, *$c*, *$d* and *$e* are parameters which are initialized by different dye names such as *RR (Rhodamine Red), RG,* etc.

```
<Simulations template="And.tret" params="$a $b $c $d $e">
  <Args values="RR RG TR AF488 LY"/>
  <Args values="RR RG TR Cy2 LY" condition="@delay>10" />
  <Args values="RR RG TR Cy3 LY" condition="@delay>5" />
  <Args values="RR RG TR Cy5 LY" condition="@delay>2" />
</Simulations>
```

2.8 Design Examples

The proposed design process can be used to find a RET network that yields a desired logic functionality. In this section, we describe the design process using two examples: a wired logic 5-input AND gate (AND5) and a wavelength sequence detector.

2.8.1 Design Example 1: AND5

Target RET Network

The functionality of AND5 can be implemented using the RET network shown in Fig. 2.8. The numbered circles and the circle labeled 'F' represent the inputs and the output, respectively. The circles labeled 'M' are mediators that act as a channel accumulating the energy of all the inputs and transferring the energy to the output chromophore. The arrows denote the preferred RET direction.

Design Rules and Metrics

The design rules for AND5 were generated using domain-specific knowledge about RET operating principles, typical chromophore attributes, and empirical results from prior lab experiments. The design rules for AND5 are as follows:

1. The Förster radius (R_0) from each of the input nodes to the mediator node should be at least 4 nm. This increases the likelihood of RET occurring from the input nodes to the mediators in the channel.
2. The R_0 from the mediator to each of the inputs should be less than 2 nm. This reduces the likelihood of RET occurring from the mediators to the input.
3. The R_0 from the mediator to the output node should be at least 4 nm. This increases the likelihood of RET occurring from the mediators to the output node.

Since AND5 is a 5-input logic gate, there are 32 different input combinations. The output node should fluoresce the most ('1') when all five input fluorophores are excited ('11111'). The metric used to assess a potential RET network for this application was defined as the ratio of the output fluorescence when all the inputs are asserted to the output fluorescence when only one of the inputs is not asserted (since all other input combinations result in a lower output, due to linearity, they are ignored). This metric is shown mathematically in 2.2, where F(abcde) represents the fluorescence of the output fluorophore when the five inputs are specified by the binary vector 'abcde'.

$$On/Off\ Ratio = \frac{F(11111)}{max(F(11110) + F(11101) + F(11011) + F(10111) + F(01111))} \tag{2.2}$$

Candidate RET Network Identification and DRC

The topology of the target RET network shown in Fig. 2.8 was loaded as a parameter to the spectral analysis tool to generate chromophore type-to-node assignments that pass the DRC. In order to explore site specification, the ideal RET network was replicated on the grid of the geometry analysis tool. Based on heuristics, the grid spacing was varied from 1 nm to 4 nm. The spectral and geometric analysis tools generated a list of approximately 65,000 candidate RET networks.

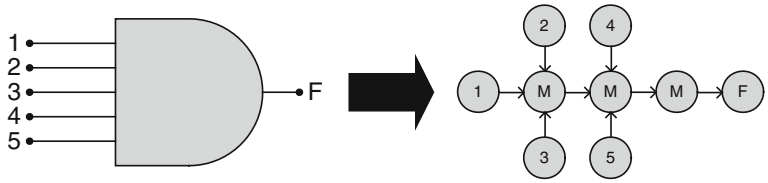


Fig. 2.8 Wired-logic 5-input AND gate can be realized on a RET network with nine nodes. Each of the numbered nodes represents an input. Node *F* is the output node. The three nodes labeled *'M'* are called the mediator nodes. The mediator nodes collect the energy from all the input nodes and transfers to node *F*. Node *F* fluoresces the most when all the inputs are excited, contributing more energy to the channel of mediators. The *arrows* represent the preferred RET transfer direction

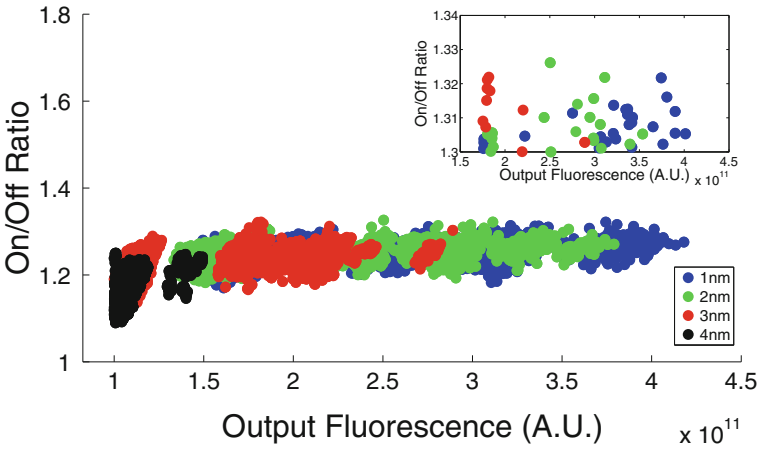


Fig. 2.9 On/off ratio versus 'on' fluorescence of unique RET networks. Each color corresponds to a different minimum resolution between two chromophores. *Inset:* 103 RET networks that meet the specifications defined using the design metrics (on/off ratio > 1.3)

Since the behavior of AND5 can be approximated by linear networks, the linear approximation tool (CTMC) was used to evaluate all candidate RET networks. Using a user-defined specification (on/off ratio must be greater than 1.3), the sample space of 65,000 candidate networks were further reduced to 103 candidate networks. Fig. 2.9 shows a graph of 'on' fluorescence of the output chromophore vs. the on/off ratio. Each dot in the figure represents a unique RET network and each color corresponds to a different minimum separation between two chromophores. The networks that have an on/off ratio greater than 1.3 make up the list of candidate networks (shown in the inset) that can be simulated in the next step of the design process.

Although the candidate networks we found here exhibit on/off ratios of only about 1.25, we have found other networks with considerably larger ratios, e.g., greater than 10^8, which are the subject of other work on non-linear RET switching devices.

Physical Simulation

The candidate RET networks were simulated using the physical simulator. The normalized output fluorescence under different input conditions is shown in Fig. 2.10. As is evident in the graph, it is necessary to set a high threshold between 'on' and 'off'. Based on the results shown in this figure, if the output threshold is set to 0.85, any fluorescence below that threshold is set to logic '0' and any fluorescence above is set to logic '1'. This ensures that this AND5 network will have an output of '1' if and only if all the inputs are on. The high threshold value is expected because the amount of energy coming into the mediator channel is directly proportional to the number of inputs that are asserted.

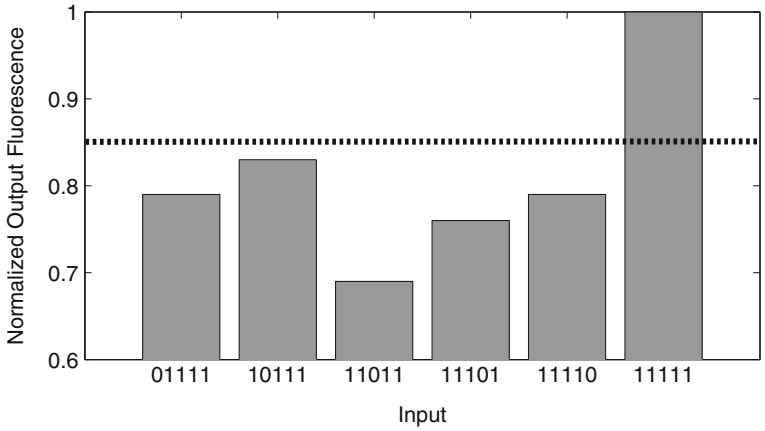


Fig. 2.10 Normalized output fluorescence of the AND5 network under six different input conditions. Five of the input conditions (everything except 11111) are expected to be the worst-case 'off' condition because in each of them, all but one input are asserted. The suggested threshold for this network is set to 0.85 (*dotted line*). This ensures that the network has an output of '1' if and only if all the inputs are asserted (when the input condition is 11111)

2.8.2 Design Example 2: Wavelength Sequence Detector

Wavelength sequence detector is an example of a non-linear RET network that operates based on acceptor saturation. A sequence detector receives delta pulses at different wavelengths (colors) of light with different arrival times and checks to see if a particular wavelength (λ_2) is preceded by another given wavelength (λ_1). In this example we assume that $\lambda_2 = 660\,\text{nm}$ (red) and $\lambda_1 = 440\,\text{nm}$ (blue) and we want them to be separated by less than 500 picoseconds. Therefore, if a red delta pulse at 660 nm arrives and in less than 500 ps, a blue delta pulse arrives at 440 nm, the detector must detect it.

Target RET Network

Since in this example we want to detect only two consecutive pulses, three nodes are sufficient to implement the detector. Two of the nodes, which we call the S and G nodes, are exclusively excited by λ_2 and λ_1 respectively and act as memory: once excited, they remain excited for a period of time. The third node is the output node and is strongly de-excited by the G node. The RET network is shown in Fig. 2.11. In this network, S absorbs the light energy and D (only) receives energy through S. The relative transfer efficiencies are chosen such that the saturation of G makes a large difference in the energy flow from S to D; because G is a strong acceptor for D and when saturated it cannot steal energy from neither S nor D. Consequently the device can act like a MOSFET switch in which saturation of G can turn the switch on.

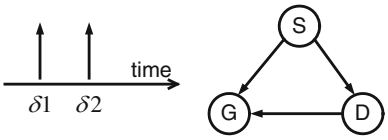


Fig. 2.11 The RET network of a sequence detector which is behaviorally similar to a normally-off switch when *G* is unexcited, the *S*-to-*G* energy flow is (almost) zero, and when G is excited, the *S*-to-*G* energy flow is high, emulating an ON switch. *Operation* Since the *D*-*G* pair is a relatively strong RET pair, *G* steals away most of the energy from *S*, either directly from it or via *D*; however, when G is pre-excited by a delta pulse ($\delta 1$) it will remain excited for a while and if the second delta pulse ($\delta 2$) arrives during this period, *S* gets excited while *G* cannot receive any more energy (already excited); therefore the energy of *S* flows to *D*

Design Rules and Metrics

In this example we want the S and G chromophores to be almost exclusively excitable by the 660 nm and 440 nm wavelengths. The G chromophore has to be an efficient donor to D and an acceptor to S. We defined the following set of spectral design rules for this example.

```
<Dyes>
  <Set name="S" Filter="$S.Abs(660) > 0.6" />
  <Set name="G" Filter="$G.Abs(440) > 0.6" />
  <Set name="D" Filter="$D.ExcPeak > 500" />
</Dyes>
<Combine>
  <Pair join="$S,$G" name="SG"
     Filter="$S.EmPeak<$G.ExcPeak && JOverlap($S,$G)>1">
  <Pair join="$SG,$D" name="SGD"
     Filter="$SG.G.EmPeak<$D.ExcPeak && JOverlap($D,$SG.G)>10">
</Combine>
```

The most important characteristic that we are looking for, in this example, is the On/Off ratio which is defined as the ratio of the excited population of D with and without saturation of G.

Candidate RET Network Identification and DRC

The defined set of spectral design rules when applied to the set of ATTO dyes (a commercial dye set with 36 dyes) generates 860 candidate type-assignments to S, G, and D. The size of the site-assignment domain on a DNA tile with eight sites is 56 (8 choose 3); since 860×56 is small enough for the physical simulation step, no further design-space pruning is required for this example.

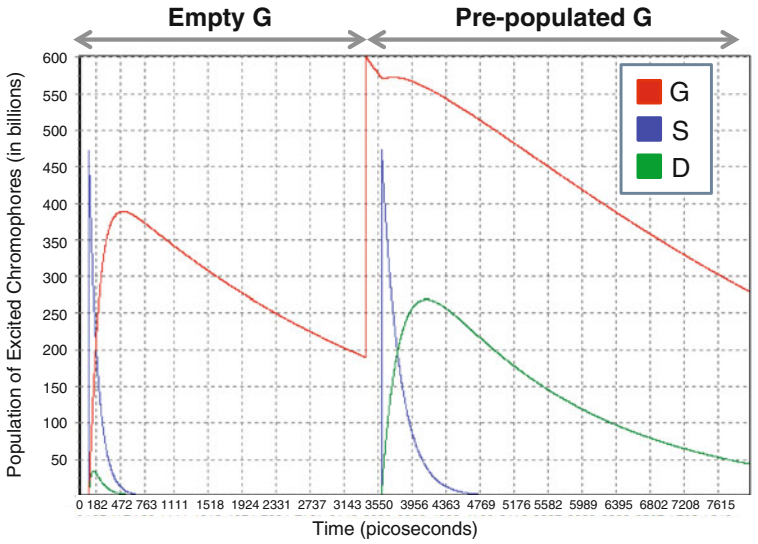


Fig. 2.12 Simulation results of the final detector network The excited population of each node is shown in the graph. *Operation:* two blue pulses are applied to the network one of which is preceded by a red pulse. As observed in the diagram, the excited-D population is ≈11 times larger after the seconds pulse

Physical Simulation

After simulating all the candidate networks we found the best On/Off ratio to be ≈11. The simulation results of the chosen network are shown in Fig. 2.12.

2.9 Evaluation

Given that all existing design methods for RET structures are *ad hoc*, this work is the first formal design process to the best of our knowledge. As such, the tool chain we describe enables much broader design space exploration than was possible previously. The fundamental difficulty that arises in RET circuit design stems from the intractability of analytical (e.g., ODE) descriptions for RET networks. Further, many such analytical methods capture only the linear response of the network despite clear evidence for the utility of non-linear behavior.

The design process we describe here also enables interactive design and custom, microenvironment (i.e., physical) simulations to capture dynamic changes during circuit operation. For example, the internal, non-linear chemical dynamics of a living cell can be incorporated in the Monte Carlo generator used in SCIMM to further

refine the prediction of a candidate network's response for *in vivo* applications such as dynamic drug delivery, sensing, and diagnostics.

2.10 Conclusion

We have described a design flow for RET circuits which can implement user-defined functionality through technology mapping to chromophores chemically assembled by DNA. The flow tracks typical EDA design methodologies with important differences in the heuristic pruning of the design space and specific metrics to evaluate candidate networks.

Despite being in the early days of development, RET circuits have the potential to bridge the gap between domains where conventional silicon technology is unable to provide computational resources. For example, RET circuits are intrinsically biologically compatible and as this emerging field matures will yield new ways to interact with biological systems through *in situ* computer-controlled processes and measurements. Thus, we have demonstrated a design approach that further enables DNA nanotechnology to expand the computational domain.

References

1. C. Pistol, C. Dwyer, A.R. Lebeck, Micro. IEEE **28**, 7 (2008)
2. H.M. Watrob, C.-P. Pan, M.D. Barkley, J. Am. Chem. Soc. **125**, 7336 (2003)
3. D. Bystranowska, B. Siejda, A. Oyhar, M. Kochman, Biophys. Chem. **170**, 1 (2012)
4. C. Pistol, C. Dwyer, Nanotechnology **18**, 125305 (2007)
5. Lifetime Data of Selected Fluorophores, (ISS) http://www.iss.com/resources/reference/data/_tables/LifetimeDataFluorophores.html
6. T. Frster, Ann. Phys. **437**, 55 (1948)
7. Fluorescence SpectraViewer, (LifeTechnologies, Invitrogen) http://www.invitrogen.com/site/us/en/home/support/Research-Tools/Fluorescence-SpectraViewer.html
8. BD Fluorescence Spectrum Viewer: A Multicolor Tool, (BD Biosciences) http://www.bdbiosciences.com/research/multicolor/spectrum/_viewer/index.jsp
9. C. Boswell, U. Utzinger, Spectra Database hosted at the University of Arizona, http://www.spectra.arizona.edu/
10. H. Zhu, V. May, B. Rder, Chem. Phys. **351**, 117 (2008)
11. J. Kowalewski, L. Mäler, in: *Nuclear Spin Relaxation in Liquids*, 1st edn. (Taylor & Francis, UK, 2006), pp. 65–95
12. J. Megow, B. Rder, A. Kulesza, V. Bonai-Kouteck, V. May, ChemPhysChem **12**, 645 (2011)
13. T. Renger, I. Trostmann, C. Theiss, M.E. Madjet, M. Richter, H. Paulsen, H.J. Eichler, A. Knorr, G. Renger, J. Phys. Chem. B **111**, 10487 (2007)
14. A. Shabani, M. Mohseni, H. Rabitz, S. Lloyd, Phys. Rev. E **86**, 11915 (2012)
15. A.A. Demidov, in: *Resonance Energy Transfer*, 1st edn. (John Wiley & Sons Ltd, New York, 1999), pp. 435–65

Chapter 3
Smart Fold Architecture: A Nano Information System Based on a Hierarchical Structure

Jun Tanida

Abstract A sophisticated model for a nano information system called the *smart fold architecture* is presented. The system is based on a hierarchical structure consisting of four essential layers and one optional layer. A systematic approach to system construction enables a variety of processing operations using state-of-the-art information technologies.

3.1 Introduction

Recent progress in biotechnology has had a strong impact on a range of scientific fields, including medical science, healthcare, biology, and so on. As is well-known, living systems are constructed from a hierarchical formation of component parts, ranging from the level of molecular systems to the level of individual organisms. From this viewpoint, biotechnology mainly explores the functions of molecular systems at the nano-scale and provides powerful tools for manipulating molecules.

Biotechnology advances have also opened the door to a new field of information science called *nano information systems*. Nano information systems are expected to reduce system footprints and decrease power consumption, and will offer increased affinity with living organisms or biological systems. Among the various methodologies that are available to us, DNA information technology is promising, owing to its maturity, accessibility, and importance for the study of living systems. A diverse range of applications of DNA, from basic logic gates to more complex optimization problems, have been proposed and demonstrated in experimental settings.

In this chapter, a sophisticated model for a nano information system, called a *smart fold architecture*, is presented. The system is based on a hierarchical architecture designed to achieve overall control of molecular systems in the living world

J. Tanida (✉)
Graduate School of Information Science and Technology, Osaka University, Osaka, Japan
e-mail: tanida@ist.osaka-u.ac.jp

M. Naruse (ed.), *Nanophotonic Information Physics*,
Nano-Optics and Nanophotonics, DOI: 10.1007/978-3-642-40224-1_3,

at the macro-scale. First, some of the fundamental concepts of DNA and DNA information technology are explained. Notable DNA information techniques relating to nano information systems are summarized briefly. After that, nano information systems are categorized according to their system architecture. Based on this background, the *smart fold architecture* is explained in terms of its concepts, features, and system operation. Following a general description, the components of the smart fold architecture are described in detail, and some promising applications are mentioned. Finally, to conclude this chapter, the implications of smart fold information systems, as well as future challenges, are summarized.

3.2 DNA Nano Information Systems

Nano information systems can be constructed effectively with state-of-the-art DNA technologies. In this section, the fundamentals of DNA information technology are explained, and some notable DNA information techniques are briefly reviewed. After that, DNA nano information systems are categorized into four classes to clarify the features of these systems.

3.2.1 Fundamental Concepts

The eccentric nature of DNA is the foundation of DNA nano information technology, and we take this as a starting point in helping to understand the concepts of DNA information techniques. How to utilize DNA in nano information technology will then be explained.

DNA

DNA has the specific feature that it forms a double helix, which is the stable structure that preserves the entirety of information related to living organisms and their heredity [1]. A simplified illustration of DNA, shown in Fig. 3.1, illustrates the mechanism of forming the double helix structure. Four kinds of bases, adenine (A), cytosine (C), guanine (G), and thymine (T), constitute a strand of DNA, where hydrogen bonds are formed between corresponding pairs, namely, (A and T) and (C and G). Therefore, an arbitrary sequence of bases has a complementary sequence that consists of bases that specifically bind to the corresponding bases in the pairs. The original and complementary sequences together form a double-strand DNA under appropriate chemical conditions, resulting in the well-known double helix structure. This phenomenon is called *hybridization*, and the characteristic of specifically binding to complementary single-strand DNA is named *Watson–Crick complementarity*.

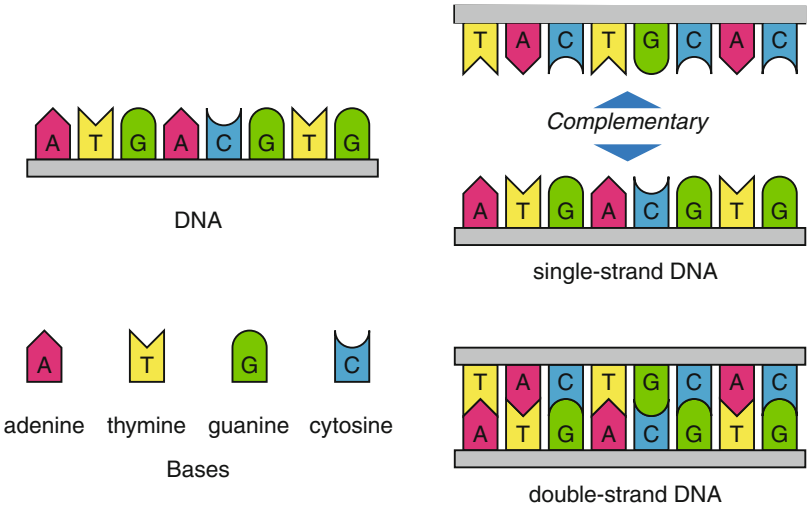


Fig. 3.1 Structure of DNA and binding between complementary strands

DNA Information Technology

After the discovery of the eccentric features of DNA, a variety of methods and techniques have been developed to manipulate DNA for various applications. Since DNA reactions are affected by temperature, pH, and the presence of catalysts, the behavior of DNA is highly controlled so that the desired manipulation can be achieved. The chemical characteristics of DNA can be estimated by a DNA design algorithm for a given DNA sequence [2]. As a result, flexible DNA manipulation can be achieved, allowing the possibility of executing information processing at the level of molecules. The huge number of molecules present in the reaction space make such a scheme promising for realizing massively parallel processing at a molecular level. For these reasons, DNA is an extremely convenient foundation for exploring processing mechanisms in the nano-world and for implementing nano information systems.

3.2.2 DNA Information Techniques

For the purpose of reviewing this attractive area, some notable work on DNA information technology will be summarized here. Because DNA information technology is quite an active research area, readers are advised to consult other literature for the latest developments [3].

Hamiltonian Path Problem

Pioneering work in the field of DNA computing involved the directed Hamiltonian path problem, which was solved by Adleman using the nature of DNA, including Watson–Crick complementarity [4]. Figure 3.2 shows a schematic diagram of the problem and the solution based on an autonomous DNA reaction. The individual fragments of single-strand DNA produce a variety of combinatorial double-strand DNAs autonomously, a process that corresponds to a search for candidate solutions. Then, a double-strand DNA satisfying a specific condition, such as the shortest non-repeating sequence, indicates the solution.

The advantages of the method are massive parallelism provided by the huge number of DNA molecules and ability to perform information processing at a molecular-level, made possible by the autonomous DNA reactions. Although this work demonstrated the potential capabilities of DNA-based computation, its implementations for large problems are not realistic due to the vast amount of DNA required as the information medium.

Automaton

Shapiro and his group have studied a finite automaton implemented with DNA and a restriction enzyme [5]. A finite automaton is an idealized mathematical model of behavior, as shown in Fig. 3.3. It consists of a finite number of states and transitions between the states according to input signals.

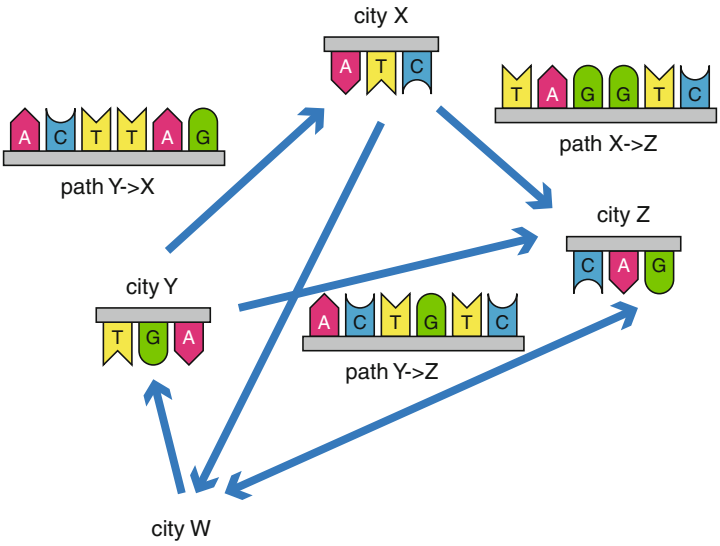


Fig. 3.2 DNA implementation of directed Hamiltonian path problem

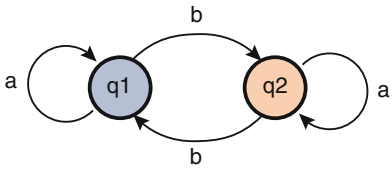


Fig. 3.3 Automaton, showing states and transitions between states

As a fundamental technique for another type of automaton, Hagiya and his group have developed a whiplash polymerase chain reaction [6]. In this method, a hairpin DNA is utilized to obtain interesting features, such as stable processing-speed for different DNA densities and the ability to accumulate the calculation status.

Logic Operations

Logic gates are the fundamental elements of digital processing. Arbitrary operations can be achieved by a combination of logical operations, so that implementation of logic gates is considered as a starting point in developing digital processing systems. Many schemes for implementing logic operations using DNA reactions have been demonstrated. For example, Winfree and his group demonstrated enzyme-free logic gates [7]. This method utilizes a reaction called *toehold-mediated strand displacement* to cascade multiple logic gates. Figure 3.4 shows a simplified implementation of an AND gate proposed by Yoshida [8]. In this method, output light is emitted only if a fluorescent dye is isolated from quenchers.

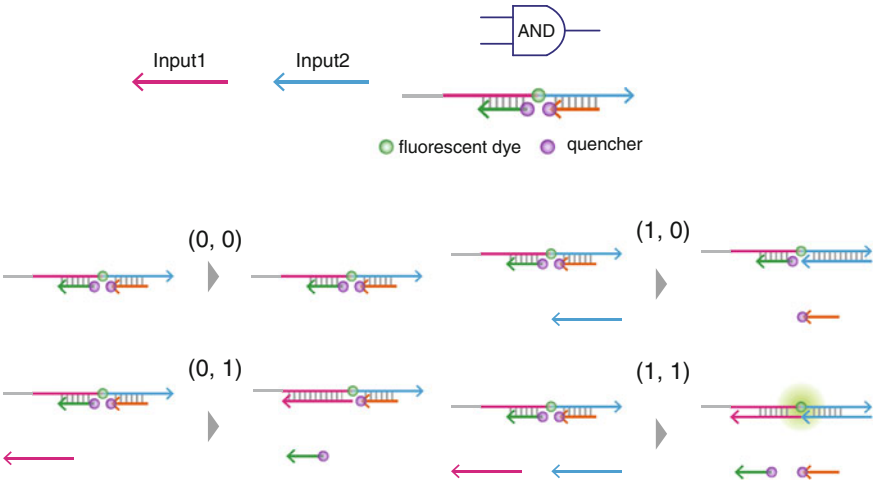


Fig. 3.4 DNA implementation of a logical AND gate

Photonic DNA Computing

The author's group has been focusing on the effective use of photonic techniques for DNA computing [9]. Photon-induced phenomena, such as localized heating within a restricted region and photoisomerization, are utilized to assist or to promote DNA reactions for the desired operations. Figure 3.5 depicts some examples of photonic DNA computing. As the temperature rises, double-strand DNAs are separated into a pair of single-strand DNAs, a process called *denaturation*. A laser beam heats a restricted region and induces DNA denaturation within that region. Photo-isomer molecules are useful for controlling the binding force between the single-strand DNAs that are bound by Watson–Crick complementarity. DNA tweezers controlled by external light irradiation were modified to perform an AND operation on the molecular inputs, forming a prototype photonic nano processor [10].

Recently, the authors presented an effective logic construction method using fluorescence resonance energy transfer (FRET)[11]. Figure 3.6 shows a conceptual diagram of the method called *scaffold DNA logic* [12]. FRET is a phenomenon in which energy transfers from one fluorescent molecule to a neighboring molecule. Single-strand DNAs having fluorescent molecules tethered to them, called *signal-transfer DNAs*, are led to specific sites on a scaffold DNA. If the signal-transfer DNAs are placed at the correct sites, an energy cascading path, via FRET, is formed, and a reporting dye tethered to the terminal DNA emits a light signal of a specific

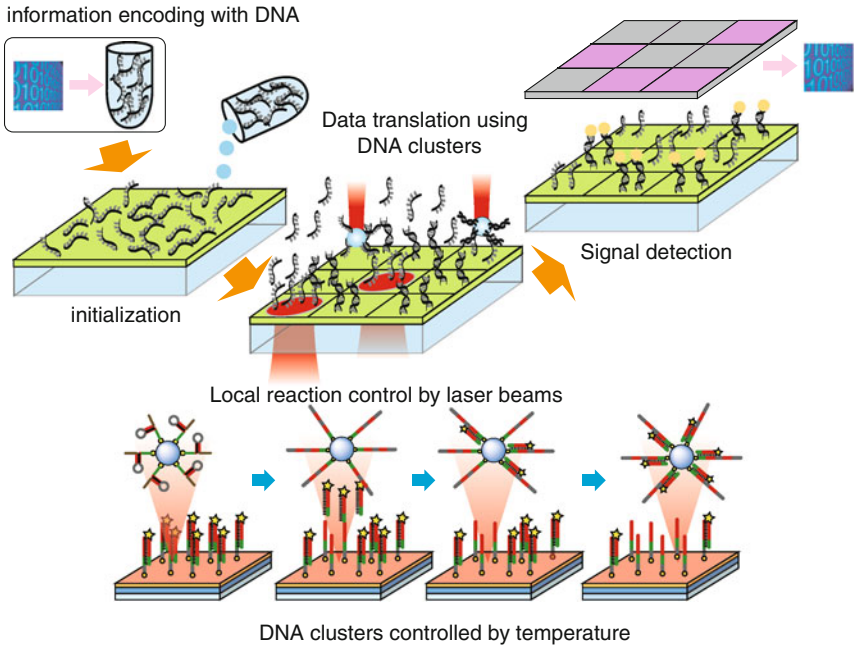


Fig. 3.5 Some examples of photonic DNA computing

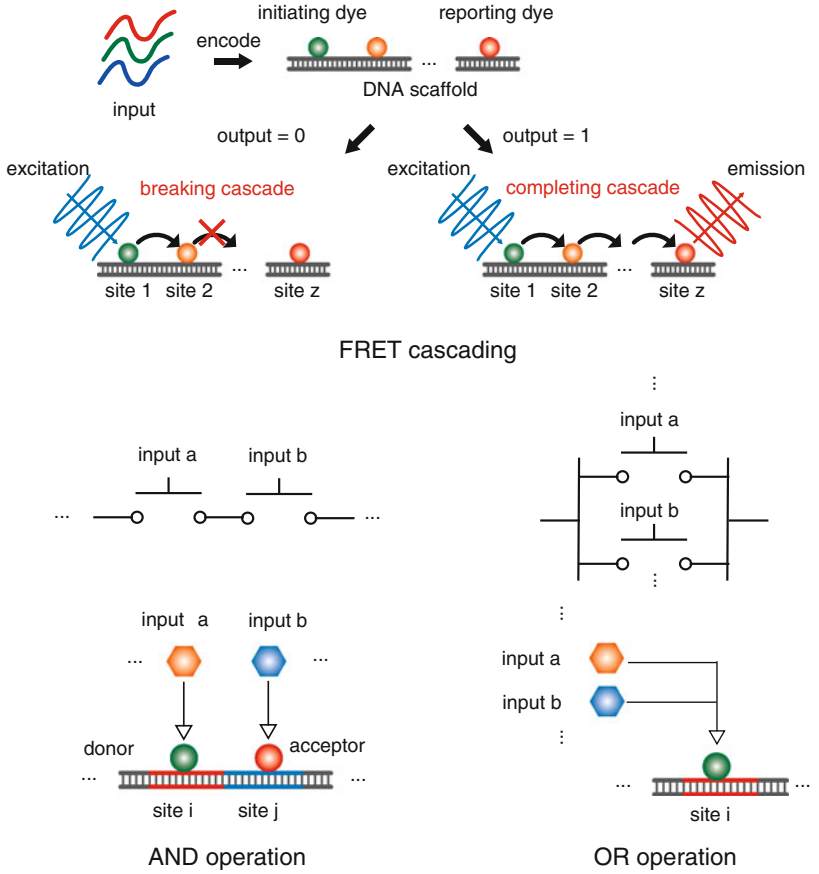


Fig. 3.6 Schematic diagram of scaffold DNA logic

wavelength. A variety of logic circuits can be designed by proper assignment of the signal-transfer DNAs and their placement sites on the scaffold DNA.

3.2.3 Architectures

Nano information systems are categorized into four classes in terms of their operating principle and interaction between the nano- and macro-scale worlds. The four classes are named *nano complete*, *nano reporting*, *nano controlled*, and *nano interactive*. Figure 3.7 summarizes these classes with typical instances.

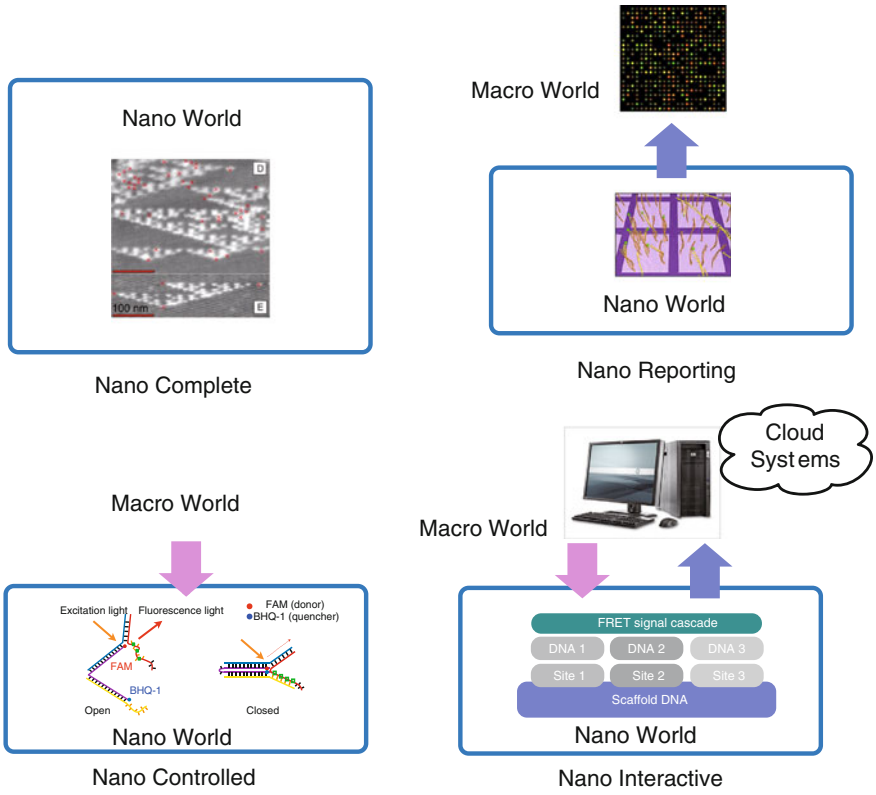


Fig. 3.7 Architectures of nano information systems

Nano Complete

Most of the traditional DNA computing schemes are designed to complete their operations within the environment of molecules, that is to say, in a so-called nano-world. In the case of typical DNA computation, the operations are embedded as a sequence of DNA intended to induce autonomous reactions based on Watson–Crick complementarity, leading to the desired result. Although the DNA preparation is done outside the nano-world, all of the processing is completed within the nano-world. From a scientific point of view, this type of processing scheme is very interesting for exploring the mechanisms of information processing hidden in the natural world and to inspire novel ideas for sophisticated information processing. Nevertheless, the current status of demonstration has remained at a primitive level, merely indicating the capabilities of the method, due to the immaturity of supporting technologies.

Nano Reporting

A DNA microarray is a typical instance of a *nano reporting* architecture [13]. Several tens of thousands of short fragments of single-strand DNA are placed on a substrate to detect their complimentary sequences in the specimen using Watson–Crick complementarity. In a preparation step, fluorescent molecules are tethered to the DNAs to be detected. Therefore, fluorescence emitted from the hybridized double-strand DNAs on the substrate indicates the existence of DNAs with the target sequence in the specimen. Fluorescence is observed through a microscope imaging system, from which the location of the emitting spots identifies the kind of detected sequence.

In this scheme, the outcome of the DNA reaction is utilized in the nano-world, whereas subsequent processing, such as signal detection and data identification, is achieved in the macro-world. This type of processing was established in the field of biotechnology and is useful in practical applications. The drawbacks are the expense and wastage involved with disposal of the devices, because they are designed for one-time use. From the viewpoint of logic algebra, a DNA microarray performs logic operations that belong to a class of combinatorial logic, meaning that the achievable processing is restricted.

Nano Controlled

The motivation of photonic DNA computing is to extend the processing capabilities of conventional DNA computing. External light signals incident from the macro-world induce the desired operations using DNA and related molecules. For the case of DNA tweeters [10], the external light signals control the open/closed states of the DNA complex in the nano-world. Although the spatial distribution of the focused light signal is much larger than the size of the DNA complex, a photoisomerization molecule tethered to the DNA receives the photon energy, and the correct operations can be achieved. The external signals are intended to control the operations inside the nano-world. In that sense, this type of processing scheme is an extension of *nano complete*, increasing the processing capability with a sort of God's hand.

Nano Interactive

The last class of nano information architecture is *nano interactive*. In this processing scheme, some processing elements work in the nano world, but they also communicate with external systems located in the macro world to execute given tasks cooperatively. A nano information system of the *smart fold architecture* is a typical instance of this class. Highly developed information systems, such as computers and cloud servers on the internet, constitute the system. Since the external computers are separated from the processing elements in the nano-world, state-of-the-art information technologies can be adopted easily. This type of nano processing scheme is suitable for constructing actual systems and applying them to practical problems.

Architecture Comparison

Figure 3.8 shows the features of the four architectures with respect to the control flow between the processors in the macro- and nano-worlds. The *nano complete* and *nano reporting* architectures employ signal transfer between the worlds just at the beginning and at the end. The *nano controlled* architecture uses one-way signal transfer to control the operations in the nano-world. In the *nano interactive* architecture, the control moves back and forth between the macro- and nano-worlds. In principle, the macro part of the processing system is effectively employed, which contributes to achieving highly sophisticated operations targeting the nano-world information.

Table 3.1 summarizes the features of the four architectures of nano information systems. As seen from the table, the *nano interactive* architecture is recognized as the most promising one. Note that communication between the macro- and nano-worlds is an important issue in making the *nano interactive* architecture a practical solution.

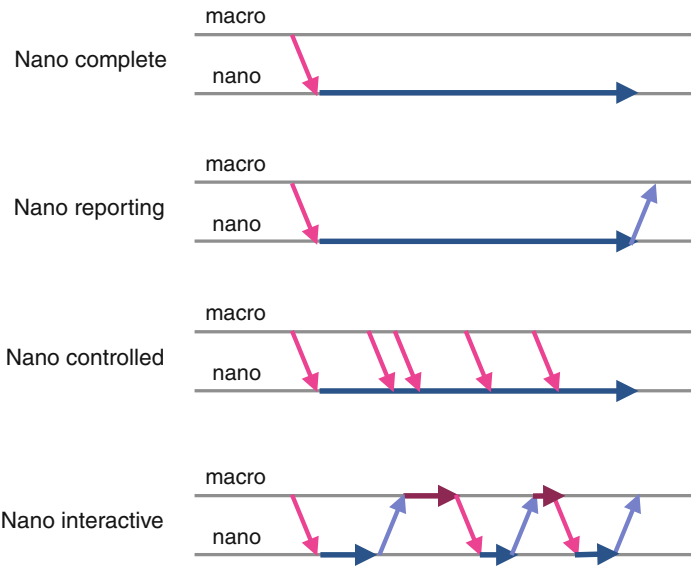


Fig. 3.8 Comparison of control flows in the four architectures

Table 3.1 Comparison of the four architectures of nano information systems

Architecture	Control mode	Communication	Tech. Maturity	Extendibility
Nano complete	Passive	None	Low	Middle
Nano reporting	Passive	Low	High	Low
Nano controlled	Active	Low	High	Low
Nano interactive	Active	High	High	High

3.3 Smart Fold Architecture

A new class of nano information system belonging to the *nano interactive* architecture, called the smart fold architecture, has been presented [14]. In this section, the details of the smart fold architecture are explained in terms of its concept, constituent layers, and system operation.

3.3.1 Concept

An effective way to understand a new proposal is to clarify the basic ideas underlying the concept. In the case of the smart fold architecture, the concept of a *smart fold* should be explained first.

What is a Fold?

The usual meaning of *fold* is a pen or enclosure in a field where livestock, especially sheep, can be kept, as shown in Fig. 3.9. Derived from this meaning, *the fold* indicates a group in a community, especially when perceived as having shared aims and values.

Smart Fold

A *smart fold* is defined as a fold possessing a degree of smartness. Smartness here means some intelligence that allows a task to be performed independently based on a decision, using internal states and external signals. A microprocessor is a typical instance of something possessing smartness. A more simple machine can also possess smartness. In our case, a DNA nano processor is regarded as a smart element, and a group of DNA nano processors corresponds to a smart fold.

Fig. 3.9 Pictures of two types of fold

Smart Fold Architecture

The *smart fold architecture* is a conceptual model of a nano information system based on the idea of a smart fold. Collective operations performed on a smart fold enables effective control of all elements belonging to the fold. In our specific implementation, optical signals are adopted to control a group of DNA nano processors using broadcast communication.

Smart Fold Information System

Figure 3.10 shows a schematic diagram of a nano information system based on the smart fold architecture. The system is based on a hierarchical structure consisting of multiple layers. The principal part of the whole system is the host-control layer. Under the host-control layer, optical signal, DNA processor, and molecular signal layers are placed. Optionally, the system can be connected to a cloud layer over the internet.

DNA Smart Fold Computing System

As an instance of a nano information system based on the smart fold architecture, a DNA smart fold computing system is considered. In the DNA smart fold computing system, a group of DNA nano processors are employed to manipulate information in the nano-world. A DNA processor is an element that processes nano scale information based on DNA information technology. An autonomous reaction of DNA is utilized to perform logical operations on bio-molecules, such as micro DNAs and RNAs.

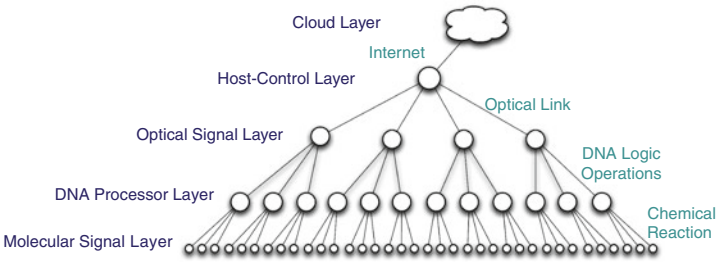


Fig. 3.10 A schematic diagram of a nano information system based on the smart fold architecture

Light signals are effectively utilized to control the group of DNA processors without caring about the individual properties of the processors.

3.3.2 Constituent Layers

A DNA smart fold computing system is assumed to consist of four essential layers and one optional layer. The four essential layers are molecular signal, DNA processor, optical signal, and host-control layers, and the optional layer is a cloud layer. Some plans for implementing such a system are described in the following.

Molecular Signal Layer

The molecular signal layer works as the front-end of the system to detect the signals coming from the molecules distributed in the nano-world and to allow interactions among them. These functions are achieved by chemical reactions between DNA strands. Figure 3.11 shows a reaction scheme for detecting specific fragments of DNA (I1–I3) as the input signals using branch migration. If the target molecules exist, the communicating molecules (C1–C3) change their forms and connect to the specific binding sites (S1–S3) of the DNA processor.

DNA Processor Layer

The DNA processor layer collects the signals from the molecular signal layer and transmits the information to the optical signal layer. The operations are performed by DNA processors based on DNA information technology. Figure 3.12 is a schematic diagram of a DNA processor implemented by a scaffold DNA technique [12]. On the scaffold DNA, logic circuits are constructed utilizing fluorescence resonant energy transfer (FRET) [11], as described above in Sect. 3.2.2. The binding sites for the communicating molecules from the molecular signal layer are placed in the FRET

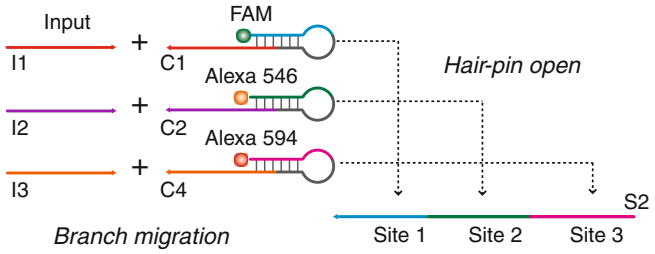


Fig. 3.11 A schematic diagram of a nano information system based on the smart fold architecture

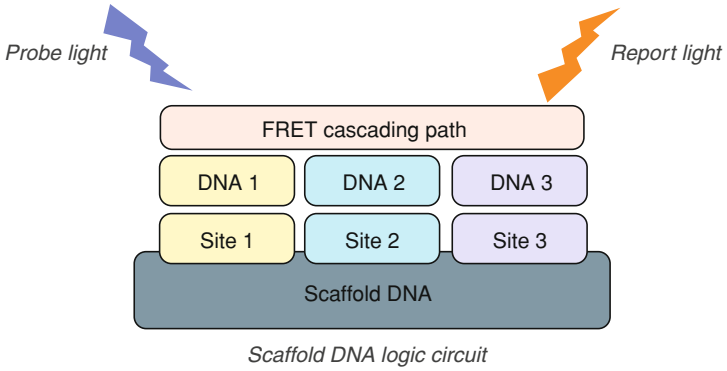


Fig. 3.12 DNA processor based on DNA scaffold technique

cascading path, and their binding states determine the function of the circuits. Probe light is supplied from the optical signal layer, and only if a specific condition for the input DNAs is satisfied, report light is emitted from the DNA processor. The operation of the DNA processor is determined by the FRET cascading path constructed on the scaffold DNA.

Optical Signal Layer

The optical signal layer functions as the interface between the nano-world and our living world. Conventional optical microscopes can be used to implement it. The spatial resolution of the optical system is restricted by the diffraction limit of light. To overcome this limitation, broadcasting and broad-gathering communication is adopted [15]. Figure 3.13 illustrates a microscope system connecting the optical signal layer to the host-control layer. A focal spot on the sample stage of the microscope is assigned to a fold of DNA processors, and the light associated with the focal spot transfers the signals to/from the DNA processors belonging to the fold. Note that the focal spot (∼sub micrometers) is much larger than the individual DNA processors (∼several nanometers), so that they cannot be individually identified in this layer.

Host-Control Layer

The host-control layer governs the whole system to execute given tasks. A standard computer system, such as a workstation, functions as this layer. A control program for the given task is executed on this layer. Necessary commands and signals are transferred to the optical signal layer and are received by the DNA processor layer. In reverse, the signals from the DNA processor layer are collected through the optical

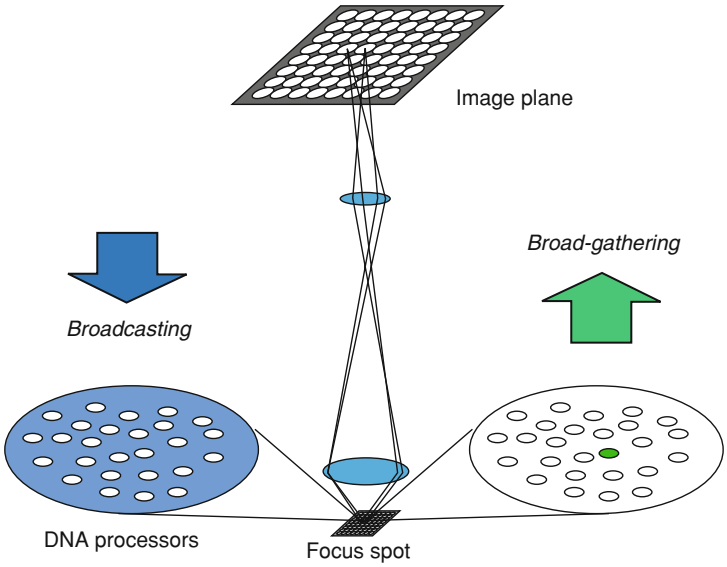


Fig. 3.13 Microscope system connecting the optical signal and host-control layers

signal layer. The collected signals are processed to retrieve the molecular signals in the host-control layer.

Cloud Layer

As an optional layer of the smart fold architecture, the system can be connected to a cloud environment through the Internet. Various services, such as computation and storage servers, can be included in the system to enhance the processing capability, as shown in Fig. 3.14.

3.3.3 System Operation

Nano information systems based on the smart fold architecture have extremely high flexibility in their operation. Both the host-control and DNA processing layers are customized to perform the desired task, and cooperative operation of these layers enables various applications. From the viewpoint of the computational model, this form of processing is regarded as a kind of client–server model, as shown in Fig. 3.15. Therefore, the processing algorithms developed for the client–server model can be utilized as the starting point for smart fold information systems. Two specific topics related to the system operation are described below.

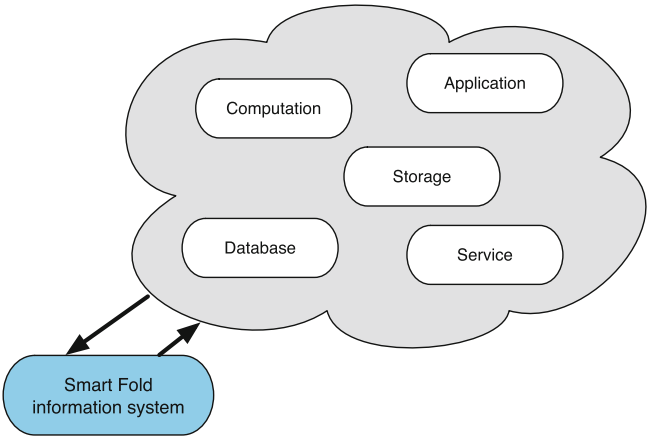


Fig. 3.14 Cloud layer

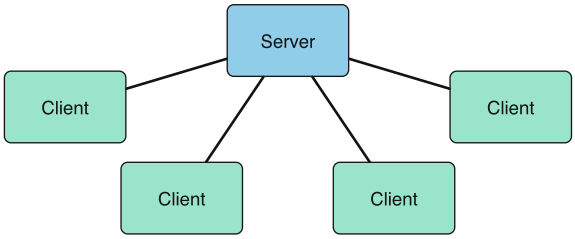


Fig. 3.15 Client–server model

Molecular Signal Reconstruction

One of the notable features of a smart fold information system is collective manipulation of the molecular signals in the nano-world. Referring to Fig. 3.10, the signals captured by the molecular signal layer are multiplexed by the DNA processing layer, transferred through the optical signal layer, and processed in the host-control layer. Although the signals detected by the host-control layer are multiplexed, appropriate encoding and decoding processing allows retrieval of the signals in the molecular signal layer.

Figure 3.16 shows a schematic diagram of molecular signal reconstruction in the DNA smart fold information system. Molecular signals captured in the molecular layer are encoded by the DNA processor and are transferred through the optical signal layer. The host-control layer detects the optical signals and executes decoding processing. Because the encoding method is intentionally designed, it can be used for the decoding process to retrieve the molecular signals. It is not expected to reconstruct complete information in the molecular signal layer, but merely to extract necessary information with the pre-designed encoding/decoding scheme.

In general, such reconstruction becomes an ill-posed inverse problem, in which the number of conditions of the observation is less than the number of signals to

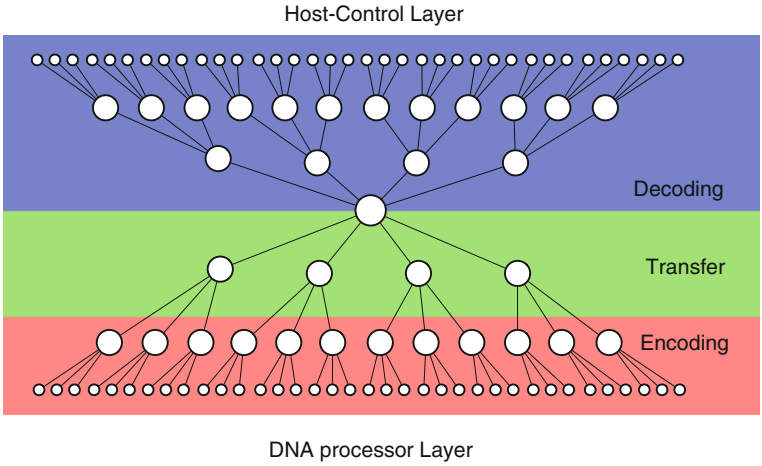


Fig. 3.16 Molecular signal reconstruction

be retrieved. The framework of compressive sensing provide a powerful solution if the target signal is sparse [16]. The author's group developed a method of retrieving multi-dimensional signals from a multiplexed observation with optical encoding and computational decoding based on compressive sensing [17]. This technique is expected to be useful in the design of encoding/decoding schemes for smart fold information systems.

Encoding and Decoding

The optical signal layer performs the communication task between the DNA processing and host-control layers. From the aspect of optical communication, the optical signal layer itself is analogous to data transfer over an optical fiber network, as shown in Fig. 3.17. As a result, effective techniques for signal multiplexing in optical networks can be adopted. For example, time-domain multiplexing (TDM), wavelength-domain multiplexing (WDM), and code-domain multiplexing (CDM) can be utilized for this purpose.

Another encoding/decoding scheme is based on asynchronous activation of DNA processors belonging to the same fold. Even if the spatial resolution of the observation optics is not sufficient to distinguish multiple DNA processors, a single DNA processor can still be observed with sufficient spacing. Asynchronous activation ensures such a sparse spacing condition and makes it possible to detect a single DNA processor. This technique is the same as that used in stochastic optical reconstruction microscopy (STORM) [18].

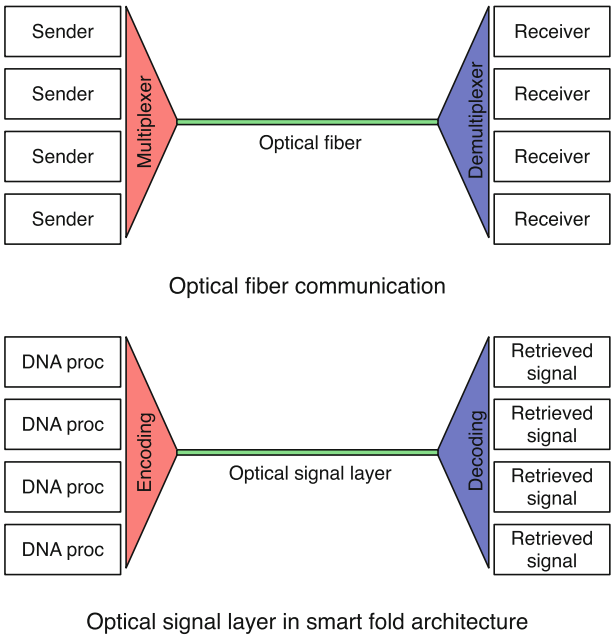


Fig. 3.17 Optical links in the optical signal layer

3.4 Applications of Smart Fold Information Systems

Smart fold information systems are expected to be employed in a wide range of applications, including functional imaging, cell monitoring, disease diagnosis, and drug delivery. In this section, some implementation ideas for promising applications of smart fold information systems are presented.

Functional Imaging

Functional imaging is a typical application of a smart fold information system for exploring molecular signals in the nano-world. The basic idea is the same as fluorescence microscopes used to visualize images of nano-scale objects. Fluorescence is utilized as a marker to indicate the position of the emitting molecule even if the optical signal is blurred by diffraction. If the emitting molecules are located sparsely enough, a single molecule can be distinguished with the help of image restoration, such as deconvolution filtering. As shown in Fig. 3.18, a limited number of DNA processors are designed to emit fluorescence, and all of the DNA processors are scanned to detect all of the molecular signals.

In the smart fold information system, a fluorescence molecule is a reporter sending the status of logical conditions in the target molecules. A notable feature of the smart

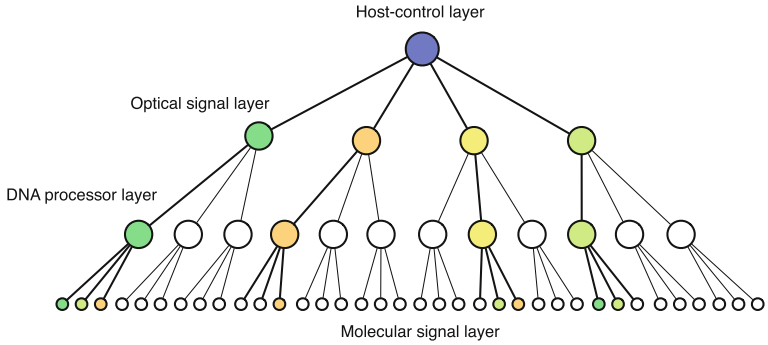


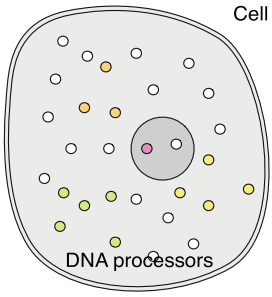
Fig. 3.18 Functional imaging by smart fold information system

fold information system is programability in designing the logical conditions in the observation. For example, if the functional activity of molecules, which depends on a specific combination of related molecules, can be described in the form of a logical condition, the reporting signal can be selectively obtained for the given condition. Compared with simple fluorescence markers, the smart fold information system is capable of handling complicated events relating to multiple molecules, so that it can visualize the functions of molecules in the nano-world.

Cell Monitoring

As an extension of functional imaging, observation of biological activities in a living cell is an attractive application. Figure 3.19 depicts the concept of cell monitoring by a smart fold information system. Appropriately designed DNA processors are injected into the target cell to detect the status of the inter-cell environment and to issue report signals. The DNA processors are distributed over the target cell, so that the spatial dependence of the specific functions can be captured. Once the function of interest is described as a logical expression, a DNA processor that executes the appropriate operation can be designed and composed.

Fig. 3.19 Cell monitoring



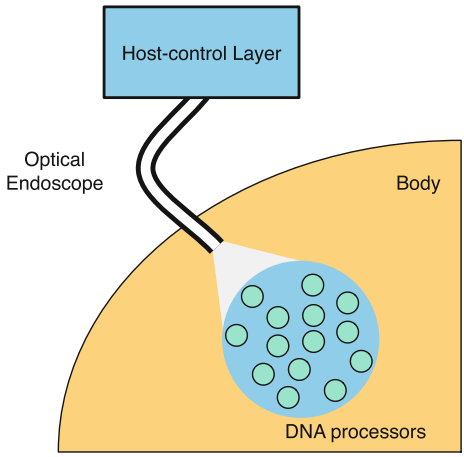


Fig. 3.20 Optical signal layer using optical endoscope

Disease Diagnosis

The cell monitoring above is directly extended to high-performance disease diagnosis. Interactivity of the smart fold information system enables us to achieve more flexible and more reliable diagnosis of cells under examination. For the case of a biopsy, the cell under examination can be placed on the stage of a microscope and easily integrated into the smart fold information system. However, for diagnosis of a living subject, an apparent issue is how to establish communication between the DNA processors distributed in the cells under examination and the host-control layer located outside the body.

One possible solution is to implement the optical signal layer with endoscope technology. As shown in Fig. 3.20 an optical link is established to connect the DNA processors and the host-control layer via an endoscope inserted near the examination area. The endoscope also provides visual information around the examination area, which provides flexibility and ensures reliable diagnosis.

Drug Delivery

Benenson and his colleague have presented the sophisticated idea of smart drugs based on DNA computing [19]. A smart drug is a kind of intelligent medicine capable of controlling its effect by sensing the internal condition of the patient, as shown in Fig. 3.21. In the demonstration, the existence of a set of DNA fragments (the condition of the patient) is detected with the same technique as used in DNA logic gates, and the stored DNA fragments (the medicine effector) are released. DNA processors can be programmed to determine the condition for releasing the medicine. Due to flexibility in the system design, the smart fold architecture is expected to increase the range of applicable cases. In addition, using an endoscope for disease diagnosis

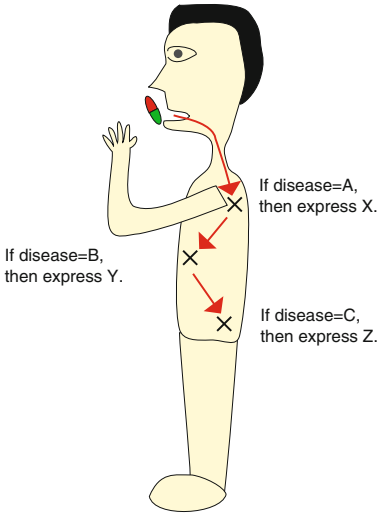


Fig. 3.21 Smart drug

in the living body can increase the effectiveness of smart drugs. This approach also combines visual inspection and medication, which is expected to increase flexibility and ensure the reliability of smart drugs.

3.5 Discussion

3.5.1 Implications of Smart Fold Computing

The smart fold architecture presents a practical solution to the problem of implementing nano information systems. Typical studies on nano information systems are categorized into two approaches in terms of the measurement scale: top-down and bottom-up ones. The former intends to shrink current micrometer-order technologies, such as semiconductor fabrication. The latter aims to build up designed systems from fragments of nanometer-order elements, such as molecules. From the viewpoint of this categorization, the smart fold architecture is considered as a hybrid approach having the advantages of both.

Another aspect of the smart fold architecture is that its implementation methodology is generalized as a system development task in the real world. System development is considered to be the task of assembling functional elements for constructing a sort of organization providing higher-level functions or services. Most machines, apparatuses, and pieces of equipment are considered as systems, and we received a lot of benefits from systems in our lives. In spite of the low functionality of the individual elements, combinations of different functions allows the emergence of

novel functionality and benefits, which is a very important advantage of building a system.

The smart fold architecture apparently follows this methodology of system construction. As introduced in the previous sections, various useful nano information techniques have been developed and are available for system construction. A hybrid approach involving a combination of top-down and bottom-up ones can pave the way to developing practical nano information systems.

3.5.2 Future Issues

Although state-of-the-art nano technology can be adopted for nano information systems based on the smart fold architecture, various issues remain before the development of practical systems becomes a reality. Some of the important issues are summarized below.

System Design

System design is the first and the most important task in developing an effective system. From the basic design to a detailed one, different levels of design tasks are required for system construction. Viewed in another way, the concept of the smart fold architecture enables us to design a nano information system with high functionality.

Operation

Once a concrete system model is designed, how to execute given problems on the system is the next important issue. Because the smart fold architecture has large flexibility in its operation, development of efficient processing algorithms is an interesting and worthwhile subject.

Implementation

Although many options exist in DNA nano information techniques, most of them are immature and still at the experimental phase. Selection of appropriate techniques holds the key to extending the capabilities of the proposed system. Well-designed demonstrations are required to show the potential capabilities of the smart fold architecture.

Optical Signal Layer

The optical signal layer is a key part in the smart fold architecture from the viewpoint of instrumentation. Microscopy for visualizing the nano-world is still a hot research subject, and many systems have been developed. Although the optical signal layer is regarded as the same category, explicit encoding and decoding processes in the DNA processor and host-control layers represent a new concept of nano-world visualization. Incorporating various types of microscopy into the smart fold architecture will serve to enlarge the choice of optical signal layers.

Applications

Killer applications are the key to establishing the position of a newly presented concept or method. Biological applications are suitable for DNA nano technology and will also be important for smart fold information systems. Intelligence and interactivity are notable features of the proposed system, so that applications that effectively use these features should be selected.

3.6 Conclusion

In this chapter, an architecture for nano information systems called the *smart fold architecture* has been presented. DNA nano technology is attractive because of its capabilities and flexibility in information processing, and a variety of methods and techniques have been developed. Among the four architecture categories mentioned in this Chapter, the *nano interactive* architecture is promising for practical system implementation, and the smart fold architecture is a representative example of this category.

The smart fold architecture is a kind of hierarchal system consisting of multiple layers, from the molecular signal layer to the host-control layer. Molecular signals in the nano-world are manipulated collectively with a conventional imaging methodology. Functional imaging, cell monitoring, disease diagnosis, and drug delivery are promising examples of smart fold information systems.

In summary, the smart fold architecture provides a sophisticated approach for implementing practical nano information systems. State-of-the-art nano information techniques are combined to achieve a high-level of functions and services following established system development methodologies. It is expected that the concept of the smart fold architecture will pave the way to developing practical nano information systems.

References

1. B. Alberts, D. Bray, K. Hopkin, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Essential Cell Biology*, 3rd edn. (Garland Science, New York, 2010), Chap. 5
2. M. Arita, S. Kobayashi, New Gener. Comput. **20**, 263 (2002)
3. A.V. Pinheiro, D. Han, W.M. Shih, H. Yan, Nat. Nanotechnol. **6**(12), 763 (2011)
4. L.M. Adleman, Science **266**, 1021 (1994)
5. Y. Benenson, T. Paz-Elizur, R. Adar, E. Keinan, Z. Livneh, E. Shapiro, Nature **414**, 430 (2001)
6. K. Sakamoto, D. Kiga, K. Komiya, H. Gouzu, S. Yokoyama, S. Ikeda, H. Sugiyama, M. Hagiya, Biosystems **52**, 81 (1999)
7. G. Seelig, D. Soloveichik, D.Y. Zhang, E. Winfree, Science **314**, 1585 (2006)
8. W. Yoshida, Y. Kobayashi, Chem. Commun. Issue **2**, 195 (2007)
9. Y. Ogura, T. Nishimura, H. Sakai, J. Tanida, in: *Sequence and Genome Analysis: Methods and Applications*, (iConcept Press, Kowloon, Hong Kong, 2010), pp. 1
10. T. Nishimura, Y. Ogura, H. Yamamoto, K. Yamamda, J. Tanida, Proc. SPIE **8102**, 810207 (2011)
11. R. Roy, S. Hohng, T. Ha, Nat. Method **5**, 507–516 (2008)
12. T. Nishimura, Y. Ogura, J. Tanida, Appl. Phys. Lett. **101**, 233703 (2012)
13. M. Schna, *Microarray Anal.* (Willey-Liss, New York, 2002)
14. J. Tanida, in: *Technical Digest of the First International Workshop on Information Physics and Computing in Nano-scale Photonics and Materials* (IPCN), (Univ. of Orléane, France, 2012), IPCN1-15
15. A.S. Tanenbaum, D.J. Wetherall, *Computer Networks*, 5th edn. (Prentice Hall, New York, 2010)
16. D.L. Donoho, IEEE Trans. Info. Theor. **52**(4), 1289 (2006)
17. R. Horisaki, K. Choi, J. Hahn, J. Tanida, D. Brady, Opt. Express **18**, 19367 (2010)
18. M.J. Rust, M. Bates, X. Zhuang, Nat. Methods **10**, 793 (2006)
19. Y. Benenson, B. Gil, U. Ben-Dor, R. Adar, E. Shapiro, Nature **429**, 423 (2004)

Chapter 4
Photonic DNA Nano-Processor: A Photonics-Based Approach to Molecular Processing Mediated by DNA

Yusuke Ogura, Takahiro Nishimura, Hirotsugu Yamamoto, Kenji Yamada and Jun Tanida

Abstract This chapter describes a photonic deoxyribonucleic acid (DNA) nano-processor, which is capable of dealing with molecules as real objects and with their associated information simultaneously in a bio-molecular environment. We present a light-activatable DNA nano-processor and DNA scaffold logic as implementation methods, and reveal their fundamental properties through experimental and analytical results.

4.1 Introduction

The importance of methods for observation and control of molecular events is increasing in a variety of fields, including life sciences and environmental technology. For example, molecular imaging, which enables us to visualize the activities of various bio-molecules, is helpful in understanding vital functions on a molecular

Y. Ogura (✉) · T. Nishimura · J. Tanida
Graduate School of Information Science and Technology, Osaka University, 1-5 Yamadaoka, Suita 565-0871, Osaka, Japan
e-mail: ogura@ist.osaka-u.ac.jp

T. Nishimura
e-mail: t-nishimura@sahs.med.osaka-u.ac.jp

J. Tanida
e-mail: tanida@ist.osaka-u.ac.jp

H. Yamamoto
Department of Optical Science and Technology,The University of Tokushima, 2-1 Minamijyousanjima-cho, Tokushima 770-8506, Japan
e-mail: yamamoto@opt.tokushima-u.ac.jp

K. Yamada
Graduate School of Medicine, Osaka University, 1-7 Yamadaoka, Suita 565-0871, Osaka, Japan
e-mail: k-yamada@sahs.med.osaka-u.ac.jp

M. Naruse (ed.), *Nanophotonic Information Physics*,
Nano-Optics and Nanophotonics, DOI: 10.1007/978-3-642-40224-1_4,

scale, and the knowledge obtained is useful for potential medical or technological applications [1, 2]. A living organism is a physical system composed of many types of matter, and it can also be considered as a sophisticated information system in which the molecules work as the information carriers. Thus, an information system capable of dealing both with matter as real objects and with the information associated with the matter is useful for obtaining valuable findings from the living organism or controlling it in an effective manner.

Processors that work in a molecular environment and deal with the information that originates from bio-molecules such as DNA, ribonucleic acid (RNA) and proteins are expected to provide an innovative methodology for analysis and control of the behavior of these bio-molecules. To realize such processors, the appropriate implementation size is essential. The size should be comparable to the size of molecules to be able to handle them directly in a molecular environment. We refer to such a small information processing apparatus with the features described above as a nano-processor.

DNA has drawn attention as a nanoscale material because of its superior features, such as Watson Crick complementarity, diversity of base sequences, and its capability for recognizing a variety of molecules [3]. For example, the self-assembly of DNA is a powerful method that can be used to fabricate two or three dimensional nanoscale structures [4–6]. The method is applicable to the arrangement of various materials with nanometer-scale precision; it is possible to form arrays of ligands or proteins, and nanoscale sensor processors based on optical resonance energy transfer [7, 8]. A variety of DNA-based nanomachines have also been demonstrated [9–11]. The features of DNA are also useful for molecular information processing. DNA computing is a promising methodology for realization of nano-processors that deal with both molecular information and real matter simultaneously. Since Adleman demonstrated the potential capabilities of DNA computing by solving a seven-city Hamiltonian path problem using DNA [12], many methods based on DNA computing have been proposed and demonstrated. Good examples of these methods include molecular automata using restriction enzymes [13] or deoxyribozymes [14], digital molecular circuits using strand displacement reactions [15], large-scale circuits using seesaw gates [16], molecular realization of a cellular automaton based on DNA self-assembly [17], and a bio-computing platform using DNAzymes [18]. Although information processing is achievable on the basis of the autonomous behavior in the DNA reactions, it usually requires complex design and precise control of the DNA reactions.

Light is another useful information carrier. Information photonics is a paradigm for the manipulation of information by the effective use of light as the information carrier. Information photonics often uses electronics, biotechnology, mechanics, and other technologies to achieve the required functionalities. Numerous interesting techniques and systems have been constructed [19], and the achievements in information photonics research are expected to be applied to the development of nanoscale information systems. However, the diffraction limit of light, by which the resolution of the light is typically restricted to around the sub-micron level for visible light, often prevents the straightforward use of light at the nanoscale. However, there are many

different types of light-matter interactions. By using these interactions, light becomes applicable to nanoscale science and technology.

By combining the ideas of DNA computing and photonics based on the effective use of light-matter interactions, a novel class of nano-processors is anticipated. For this purpose, we are studying a photonic DNA nano-processor with the functions of sensing molecular information, processing this information, and then actuating a physical action as an output [20, 21]. This is a new concept for measurement and control of a molecular system by cooperative use of DNA as the material for construction of the system and light as the interfacing carrier between the molecular system and the macro-world.

In this chapter, recent research achievements in the photonic DNA nano-processor field are described. We show the features of the photonic DNA nano-processor that acts as a mediator and performs an intermediate role between a molecular system and the macro-world. A light-activatable nano-processor and DNA scaffold logic are introduced as implementation techniques, and their properties are described.

4.2 Photonic DNA Nano-Processor

4.2.1 Concept

Photonic DNA nano-processors are nanoscale information-processing instruments that work in wet environments, and they are expected to be applied to, for example, bio-molecular systems in living organisms. The concept of the photonic DNA nano-processor is illustrated in Fig. 4.1. Photonic DNA nano-processors consist of DNA strands with different sequences, and the implementation size can therefore be of the order of nanometers. Ideally, photonic DNA nano-processors are capable of sensing or identifying molecular information, computing that information, and performing a physical action that depends on the results of the computation.

DNA molecules can interact with other biomolecules such as other DNAs, RNAs, and proteins. Therefore, the photonic DNA nano-processor is compatible with bio-molecular systems and is thus capable of capturing information related to the bio-molecules or indeed of controlling the bio-molecules. From this viewpoint, photonic DNA nano-processors are therefore not a straightforward downsizing of the conventional electronic processors that usually support our activities in the current information society. They provide novel functionalities, including on-site detection and action, because of their ability to access the bio-molecular systems directly. This property is important because it enables us to consider a molecular system not only as a target system to be measured or controlled by using the nano-processors, but also as an engineered system to be used for a particular purpose.

Matter that is responsive to light is incorporated into part of the DNA molecules of the photonic DNA nano-processor to construct a path for communication between the nano-processor and the macro-world via photonic signals. For example, the photonic

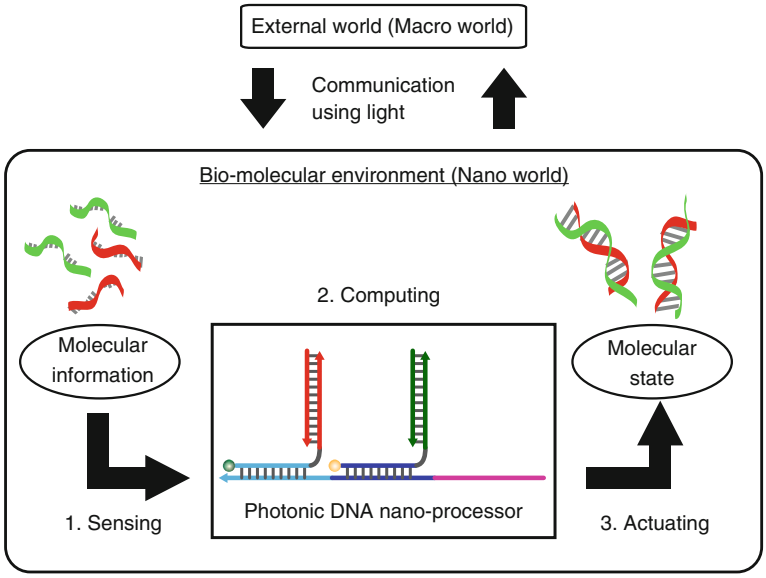


Fig. 4.1 Concept of a photonic DNA nano-processor

signals can be used to transmit control commands or external information. They can also be used to extract information from the nano-world and bring it to the macro-world. Incorporation of this photo-responsive mechanism provides a variety of functions. For example, the nano-processors can be activated at desired times. This enables stepwise progress in the processor's behavior, cooperative operation between multiple nano-processors by synchronization, and adaptive control to reflect the status of the target system. The environment of the target system is not destroyed because the photonic signals are transmitted remotely and non-invasively. In addition, the possibility of activating the nano-processors in localized volumes in parallel by distributing light into micrometer-ordered volumes is a major advantage of the use of light, because it offers position-dependent regulation of the molecular system.

4.2.2 Photonic DNA Nano-Processor as Mediator

Photonic DNA nano-processors can communicate with both the macro-world through the medium of light and the nano-world through DNA. From this viewpoint, the photonic DNA nano-processors can be considered as intelligent mediators that act as an intermediate layer between the nano-world and the macro-world.

Figure 4.2 shows a schematic diagram for acquisition of molecular information from a molecular system to the macro-world without and with a photonic DNA nano-processor as a mediator. By the conventional method (i.e. without the

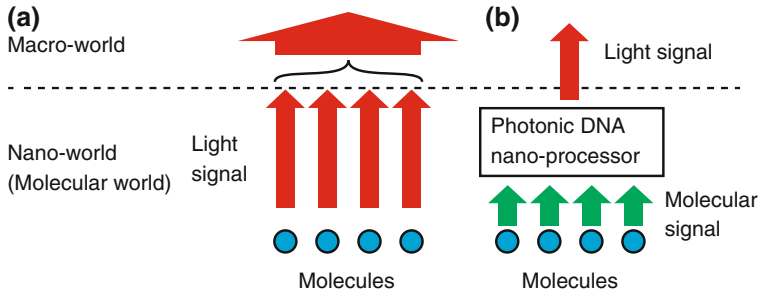


Fig. 4.2 Comparison of the information flow when molecular information is transferred from the nano-world to the macro-world, **a** without and **b** with a photonic DNA nano-processor

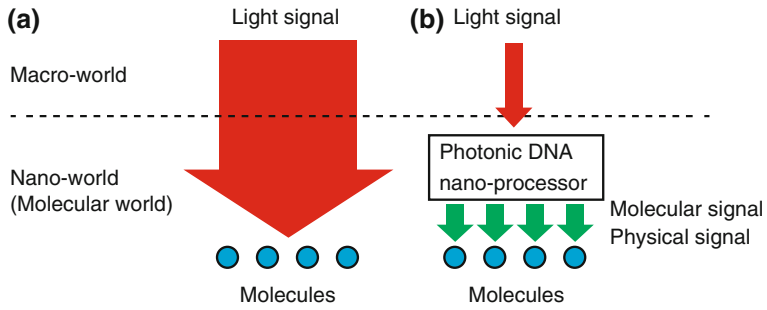


Fig. 4.3 Comparison of the information flow when a molecular system is controlled by using light, **a** without and **b** with a photonic DNA nano-processor

nano-processor), the information related to individual molecules is transferred in a straightforward way, so that a wide information bandwidth is required (Fig. 4.2a). For example, if the resolution of the imaging system is not high enough, i.e., the information bandwidth is not large enough, then the molecular information can become unexpectedly confused. In contrast, as shown in Fig. 4.2b, by using the photonic DNA nano-processor, preprocessing of the molecular information can be performed at the nanoscale to modify the information into the desired form. Logical operations and correlation operations are good examples of the preprocessing steps that can effectively suppress the amount of information. Suppression of the information at the nanoscale leads to a relaxation of the performance requirements of the apparatus used to acquire the necessary information.

Figure 4.3 shows a comparison of the methods for controlling the molecules using light. By the conventional method (Fig. 4.3a), the flexibility in the manipulation of molecules using light is restricted by the scale gap between the molecules and the light. This means that it is difficult to control individual molecules independently within the size of the light distribution. Also, direct manipulation of the molecules requires interaction between the light and the target molecules themselves. However, as shown in Fig. 4.3b, the use of a photonic DNA nano-processor offers a way to

manipulate molecules that are unresponsive to light. By using light as the trigger of the photonic DNA nano-processor, simple photonic signaling to the molecular system induces well-regulated behavior based on the capabilities of the DNA in autonomous reactions. This can be considered, from the information viewpoint, as a small amount of information carried by the light being expanded into sufficient information to control the molecular system by the photonic DNA nano-processor. Thus, despite the simplicity of the signaling, the complex behavior of the molecules can be controlled based on the autonomous reactions.

4.2.3 Types of Light Usage

The photonic DNA nano-processor has sensing, computing, and actuation functions, and photonics-based techniques can be used in these individual functions. We have considered two types of light usage in the photonic DNA nano-processor, as shown in Fig. 4.4. Light is used for activation control of the sensing function in usage type (i). In contrast, a photonic process, nanoscale photonic signal transmission, is applied to achieve the computing function in usage type (ii).

In usage type (i), the final structure of the photonic DNA nano-processor is constructed in the preliminary stage by self-assembly of DNA molecules, and then it is placed into the molecular environment. The activation of the sensing function of the nano-processor is controlled using a light signal through the response of photoresponsive molecules equipped as part of the nano-processor. This scheme offers control of the nano-processor in a specified volume of micrometer-order at the desired times because generation of arbitrary spatio-temporal patterns of the controlling light is possible. Computation is performed based on the autonomous behavior of the DNA after activation of the sensing process using the light. This photonic control ability is applicable to a variety of computation schemes based on DNA computing. It will be possible to select the function to be activated by using multiple wavelengths for the control light.

In usage type (ii), elemental DNA molecules acting as the building blocks of a photonic DNA nano-processor are inserted into a molecular system. Self-assembly of the elemental DNA molecules is used as a part of the computation process. Photoresponsive molecules are attached to some of the elemental DNA molecules, and the state of the target molecular system is encoded into the arrangement of the photoresponsive molecules. Photonic signal transmission is used to assess the molecular environment. Based on this type of usage, large-scale and high-throughput processing can be achieved, because simple DNA reactions can be used and the process proceeds in parallel by replacing the DNA reactions into a photonic process as part of the computation.

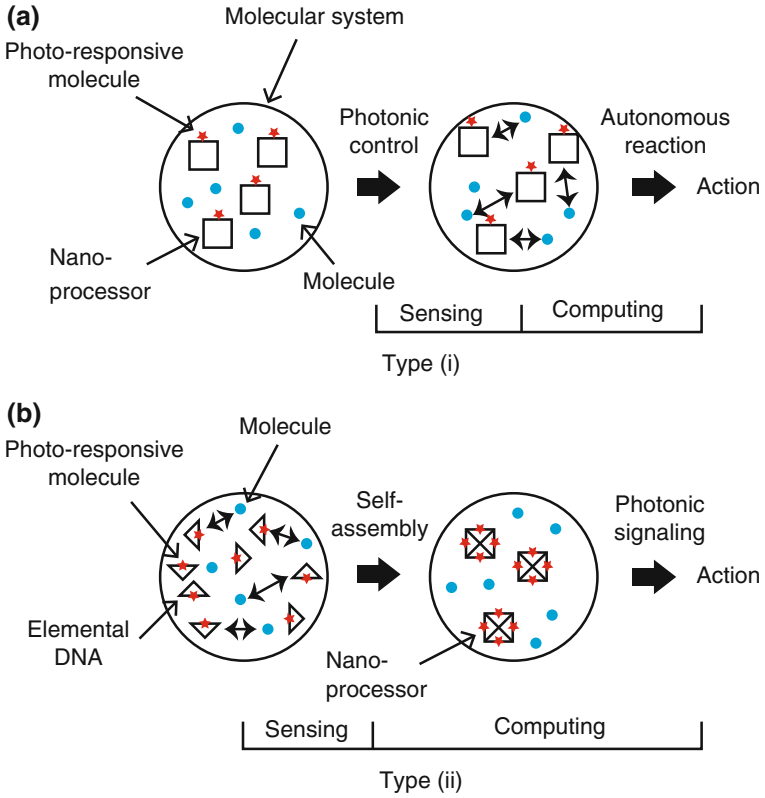


Fig. 4.4 Two types of light usage. **a** Type (*i*): the light is used to activate the sensing function. **b** Type (*ii*): a photonic signal is used as part of the computation process

4.3 Light-Activatable DNA Nano-Processor

4.3.1 Activation of Sensing Function Using Light

A photonic DNA nano-processor with a light-activatable sensing function was constructed as a prototype of usage type (i). The light-activatable DNA nano-processor senses two particular sequences of the DNA strands, decides whether the sensed DNA contains a target sequence, and transforms as an action. One distinctive feature is that the sensing function can be activated or inactivated by using photonic signals. For this purpose, the DNA tethered with a photo-isomerization molecule, azobenzene, was used [22]. The azobenzene-tethered DNA binds with its complementary DNA to form a double-stranded DNA when the azobenzene is in the trans-form upon visible light irradiation. This double-stranded DNA is separated into single-stranded

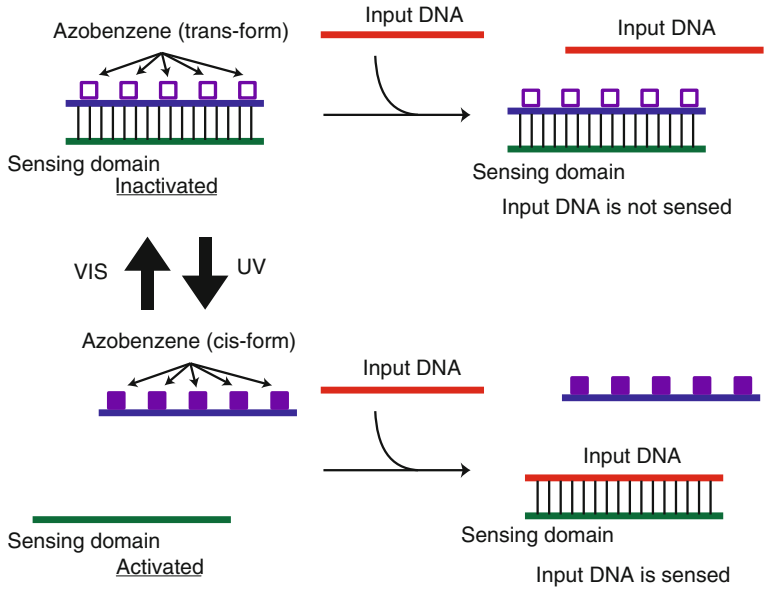


Fig. 4.5 Method for controlling the activation of the sensing function. Upon UV light irradiation, the sensor is activated, and upon visible light irradiation, the sensor is inactivated

DNAs when the azobenzene is in the cis-form upon ultraviolet (UV) light irradiation. This property enables the activation of the sensing function and also its inactivation.

Figure 4.5 shows the basic activation control scheme. The sensing domain binds with its complementary sequence with azobenzene upon visible light irradiation. In this case, the sensing function is inactivated, because the sensing domain is shielded to prevent it from binding with other DNA fragments. However, the sensing domain is exposed upon UV light irradiation. As a result, the sensing function is activated because the sensing domain is ready to bind with other DNA fragments. In this case, the input DNA can be sensed.

4.3.2 Scheme

The behavior of the light-activatable nano-processor is shown in Fig. 4.6. The nano-processor forms a tweezer-like structure with two hairpin DNAs which incorporate azobenzene molecules for photonic control. DNA tweezers were first demonstrated by Yurke et al. [9] as a DNA nanomachine by using a strand displacement reaction. Photonically-controlled DNA tweezers have also been demonstrated as a DNA nanomachine [23, 24]. However, the nano-processor with activation control is substantially different from the DNA tweezers described above.

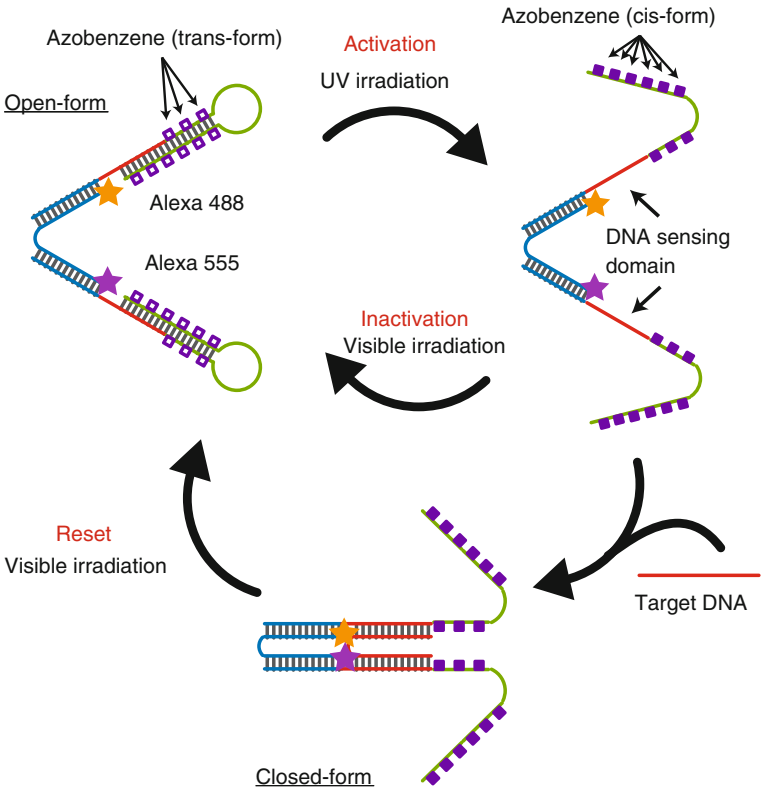


Fig. 4.6 Reaction scheme of light-activatable DNA nano-processor

In the sensing function, the nano-processor binds to a particular DNA fragment. The hairpin DNAs located at the individual edges contain sensing domains with sequences that are different from each other. This means that the individual hairpin DNAs can sense sequences independently. The hairpin DNAs open upon UV light irradiation, and the sensing domains are exposed to start the sensing process. The hairpin DNAs close upon visible light irradiation to inactivate the sensing function.

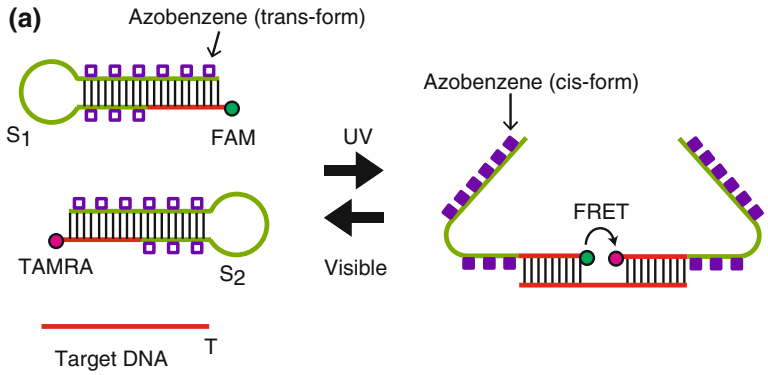
In the computing function, the nano-processor judges whether a DNA fragment binds to both sensing domains on the individual arms of the tweezers. Then, for the actuating function, the nano-processor closes its structure or maintains the open form, depending on the result. If the sensed DNA fragment contains the target sequence, then it is captured by both arms. This means that "the result is YES," and the nano-processor assumes the closed form. However, if the fragment does not contain the target sequence, it is captured by only one or neither arm. This means that "the result is NO," and the structure maintains the open form. Note that the nano-processor changes from the open form to the closed form by binding with the target DNA

only during the periods of activation. During the periods of inactivation, the nano-processor does not change its form, regardless of the existence of the target DNA.

Another interesting function is the possibility of reset. By irradiation with visible light, the target DNA captured during the period of activation is released because the hairpin DNAs close, and then the nano-processor transits to the open form (the initial form). Photo-activation of DNA probes was previously demonstrated by Wang et al. [25], but repeated sensing was impossible because of the lack of a reset function. The reset function enables the photonic DNA nano-processors to be used repeatedly.

4.3.3 Experiments

Before demonstration of the entire nano-processor, we investigated the behavior of a sensor consisting of two hairpin DNAs tethered with azobenzenes to clarify the properties of the sensing function [26]. The scheme and the DNA sequences used are shown in Fig. 4.7. One of the fluorescent dyes, 6-FAM (excitation/emission: 494/517 nm) and TAMRA (excitation/emission: 565/580 nm), is attached to each of the individual hairpin DNAs. These dyes form a fluorescence resonance energy transfer (FRET) pair. Under UV light irradiation, sensing is activated because the sensing domains are exposed as described in Sect. 4.3.1. When the target DNA exists during the period of activation, the target binds to both hairpin DNAs, and the fluorescent



(b)

S_1: 5' FAM-ATCTGAACTAACGCTxTTxAAxGATCTCTxCTxTAxAAxGCxG--TxTAxGTxTCA 3'

S_2: 5' CACxGTxACxATxGCTxTTxAAxGATCTCTxCTxTAxAAxGCATGTAC--GTGCTA-TAMRA 3'

T: 5' CGTTAGTTCAGATATCATAGCACGTACATG 3'

Fig. 4.7 **a** Scheme of the light-activatable sensor using two hairpin DNAs with azobenzene. **b** Sequences and modifications of the DNA strands used. The symbol x indicates the position of the azobenzene

dyes of the FRET pair are placed closely together, as shown in Fig. 4.7. In this case, the FRET occurs. By irradiation with visible light, the sensor returns to its initial form to be reset. The sensing results can be obtained by measuring the fluorescence signals.

We used a spectrofluorometer (JASCO CORPORATION, FP-6200) for switching activation/inactivation and for measurement of the fluorescence intensity. A xenon lamp was used as the light source. The wavelengths of the light used were 340 ± 10 nm (5 mW/cm^2) for activation and 440 ± 10 nm (7 mW/cm^2) for inactivation. The 494 nm wavelength light was used for excitation of FAM, and its fluorescence intensity was measured at 517 nm (emission wavelength of FAM). Irradiations for activation and inactivation were performed at 80 °C for two minutes, because the efficiency of the photoisomerization of azobenzene between the cis-form and the trans-form is high at high temperatures. The fluorescence intensities were measured at 25 °C.

The first experiment was performed to measure the dependence of the FRET ratio on the static concentration of the target DNA. The FRET ratio is defined as $1 - F/F_0$, where F is the fluorescence intensity of the donor of the FRET pair after sensing and F_0 is that before the sensing process. Figure 4.8 shows the relationship between the FRET ratio after activation and the concentration of the target DNA when the concentration at the sensor is 0.5 μM. The FRET ratio is approximately proportional to the target concentration. This result indicates that it is possible for this sensor to measure the target concentration.

The second experiment was conducted to investigate the dependence on the concentration of the target DNA when the concentration varied. Figure 4.9 shows the changes in the FRET ratio when the target DNA increased or decreased.

The target concentration was changed from 0.1 μM to 0.3 μM by adding the target DNA in the experiment of Fig. 4.9a, and it was changed from 0.4 μM to 0.2 μM by adding 0.2 μM of the complimentary DNA of the target shown in the experiment of Fig. 4.9b. The sensor concentration was 0.5 μM. These results demonstrate that the FRET ratio changed according to the concentration at the activation time. The FRET ratio values are almost in agreement with the results of Fig. 4.8. The sensor is suitable for reporting the concentration of the target DNA at the desired timings, regardless of the concentration at the previous measurement.

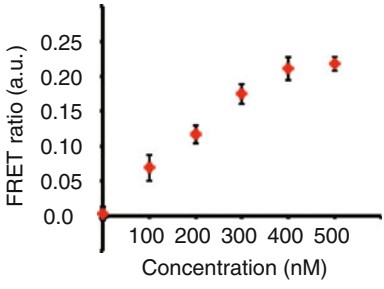


Fig. 4.8 Dependence of the FRET ratio on the input DNA concentration

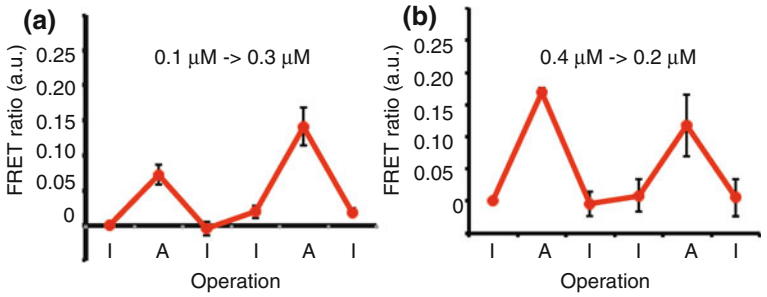


Fig. 4.9 The FRET ratio measured when the input DNA concentration changes. **a** The input DNA increases from 0.1 to 0.3 μM. **b** The input DNA decreases from 0.4 to 0.2 μM. A: Activated (UV light), I: Inactivated (Visible light)

Next, we constructed the light-activatable nano-processor shown in Fig. 4.6 to demonstrate its function. One of the pair of fluorescent dyes, Alexa 488 (excitation/emission: 495/520 nm) and Alexa 555 (excitation/emission: 550/570 nm), was attached to each of the nano-processor's arms to act as a donor and an acceptor of a FRET pair to report on the nano-processor's form. The fluorescence intensity of Alexa 488 is high when the nano-processor is in the open form, but the intensity becomes low when the nano-processor is in the closed form. The sequences of and modifications for the light-activatable nano-processor, which consists of three DNA strands, are summarized in Fig. 4.10. The FRET ratio was measured under three different conditions as follows: (i) a solution containing 5-μM target DNA and the 1-μM nano-processor was activated and inactivated three times in succession; (ii) the solution containing only the 1-μM nano-processor (no target DNA) was activated and inactivated once, (iii) the solution containing 5-μM target DNA and the 1-μM nano-processor was left without any irradiation. The irradiation processes for activation and inactivation were performed at 60 °C to maintain the structures of the light-activatable nano-processors. The fluorescence intensities were measured at 37 °C.

The FRET ratio measured after the individual operations is shown in Fig. 4.11. The FRET ratio increased upon UV light irradiation (activation) and decreased

Hairpin 1: 5' Alexa488-ACAGTTTGTCCTGGGCATCAGCTGCCGTGCTTTGCG-
-Alexa555 3'
Hairpin 2: 5' GCCCAGGACAAACTGTCTACTACCTCACxCTxAGxCTCTTCTG-
-CxTAxGGxTGxAGxGTxAG 3'
Hinge: 5' GTxATxGGxTTxCGxTTxAGTCTTCTCxTAxACxGAACCATACAACC-
-GCAAAGCACGGCAGC 3'
Target: 5' TGAGGTAGTAGGTTGTATGGTT 3'

Fig. 4.10 Sequences and modifications for the light-activatable nano-processor. The symbol x indicates the position of the azobenzene

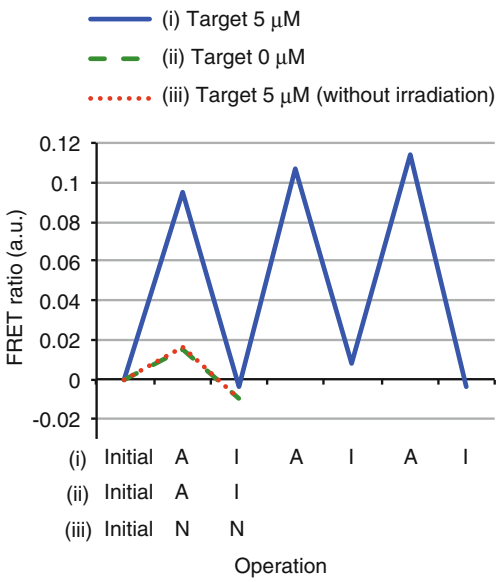


Fig. 4.11 Variance of the FRET ratio during repeated activation and inactivation operations of light-activatable DNA nano-processor. *A* Activated (UV light), *I* Inactivated (visible light), *N* Neutral (no light)

upon visible light irradiation (inactivation) for condition (i). However, the FRET ratio changed little for conditions (ii) and (iii). This result indicates that activation/inactivation of the sensing function is controlled via the external photonic signals, and the nano-processor transforms under the existence of the target DNA during the activation periods. Note that the measured FRET ratios during the activation periods are almost the same value for any number of repetitions. This suggests that stable behavior is obtained during the repetition process. Also, the FRET ratio decreased to approximately 0 after inactivation. This behavior is considered to be the result of the nano-processor releasing the previously detected DNA and returning to its initial form. The experimental results demonstrate that the nano-processor transforms, depending on whether or not the target DNA exists, and that it can be reset to work repeatedly by synchronization of the activation control using light.

4.4 DNA Scaffold Logic

4.4.1 Fundamental Scheme of Logic Operation

The DNA scaffold logic [27], which we have proposed, is a typical example of type (ii) usage of the light. The nano-processor based on DNA scaffold logic is capable of executing a logical operation with the DNA inputs and producing an output as a fluorescence signal. In this scheme, a DNA scaffold, on which the photonic process is performed, is prepared. The DNA scaffold sequence depends on the given logic

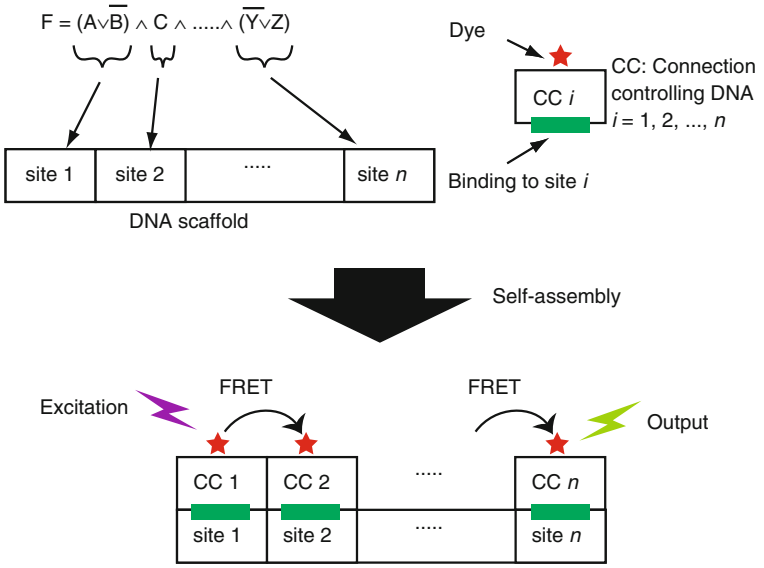


Fig. 4.12 Logic operation by DNA scaffold logic. Fluorescent dyes are placed on a scaffold according to the input DNA. The result is reported by the FRET cascades

function. Connection-controlling DNAs with a fluorescent dye are also used. The connection-controlling DNAs interact with the input DNAs, and place the fluorescent dye on the DNA scaffold or remove it from the scaffold. As a result, the target molecular system status is encoded into an arrangement of fluorescent dyes on the DNA scaffold. This process is performed based on the self-assembly of DNA.

To explain the process in more detail, let us consider that a given logic function is in the conjunctive normal form with n clauses, as shown in Fig. 4.12. A single-stranded DNA is used as a DNA scaffold. The DNA scaffold contains n sites (site 1, site 2,..., site n), and the different connection-controlling DNAs can bind at individual sites. The individual sites correspond to the different clauses. A connection-controlling DNA with a fluorescent dye binds to a particular site when the corresponding clause is TRUE for the input DNA. If the fluorescent dyes bind to site i and site $i + 1$, the energy of the excited dye on site i then transfers to another dye on site $i + 1$. Thus, a FRET path from site 1 to site n is constructed on the scaffold, but only if the input satisfies all of the clauses; i.e. that the given logic function is TRUE. As a result, the dye on site n produces a fluorescent signal by excitation of the dye on site 1. If one or more clauses are not satisfied, then the FRET path is broken on the way from site 1 to site n. As a result, the dye on site n produces no fluorescent signal to show that the logic function is FALSE.

4.4.2 Features

The DNA scaffold logic has distinctive features as nano-processors. Figure 4.13 shows a comparison between the conventional DNA logic and DNA scaffold logic. In conventional DNA logic, a number of the DNA complexes are used as computing elements, and cascaded reactions between the elements via molecular signals are used for computation. Thus, the computation generally proceeds in a sequential manner. In contrast, in DNA scaffold logic, the molecular signals are converted to an arrangement of fluorescent dyes, and part of the computation is performed by FRET signaling on the single DNA scaffolds. The FRET signaling path is constructed by placing connection-controlling DNAs with fluorescent dyes at different sites on the scaffold. The placement process proceeds in parallel and independently. This scheme no longer requires the cascaded reactions via the molecular signal. Therefore, the operation is fast, and complicated logic operations can be executed with a few additional time periods for the operation. Self-assembly is used to arrange the fluorescent dyes as part of the computation, and therefore the operation is expected to be powerful. Also, the error caused by the DNA reaction does not propagate to subsequent reactions because the reactions are designed in an independent manner. Signal transmission is performed using a FRET cascade, which is not a diffusion process of molecules, and the signal transmission efficiency is high.

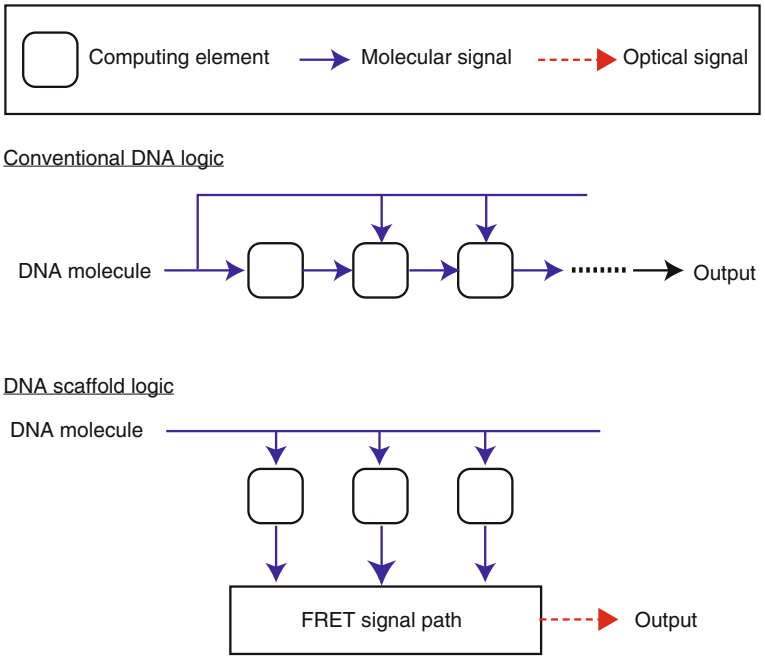


Fig. 4.13 Comparison of logic operations by conventional DNA logic and by DNA scaffold logic

4.4.3 Experiments

We designed two reactions for connection-controlling DNAs to arrange fluorescent dyes on a DNA scaffold. The first is a reaction using a connecting DNA to realize the function of a YES gate. The second is a reaction using a disconnecting DNA to realize the function of a NOT gate. The schemes of these reactions are shown in Fig. 4.14. For the YES gate, a hairpin DNA with a fluorescent dye is used as the connecting DNA. When an input DNA binds to the connecting DNA through a strand displacement reaction, the domain to be bound to the scaffold is exposed, and then the connecting DNA binds to the scaffold to carry the dye on the scaffold. For the NOT gate, a disconnecting DNA with a dye is bound on the scaffold; i.e. the fluorescent dye is arranged on the scaffold previously. When an input DNA exists, it binds to the disconnecting DNA and removes the dye from the scaffold. In contrast, the dye remains on the scaffold when the input DNA does not exist. Note that a variety of sequences can be used as the sequence for the domain for recognition of the input DNA or for binding to the scaffold. These schemes are therefore useful for dealing with large numbers of inputs and complicated logic functions.

Several logic functions were tested to demonstrate DNA scaffold logic. Figure 4.15a shows the scheme and the results of the logic function $I_1 \wedge I_2$ used to demonstrate the AND operation. The sequences and modifications that were used

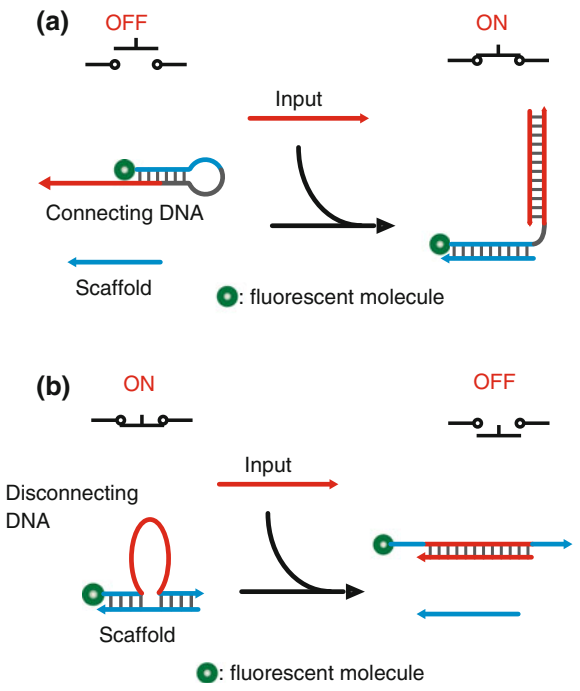


Fig. 4.14 Reaction schemes for **a** a connecting DNA and **b** a disconnecting DNA. When using the connecting DNA, the dye is placed on the scaffold when an input DNA exists. When using the disconnecting DNA, the dye is put on the scaffold when the input DNA does not exist

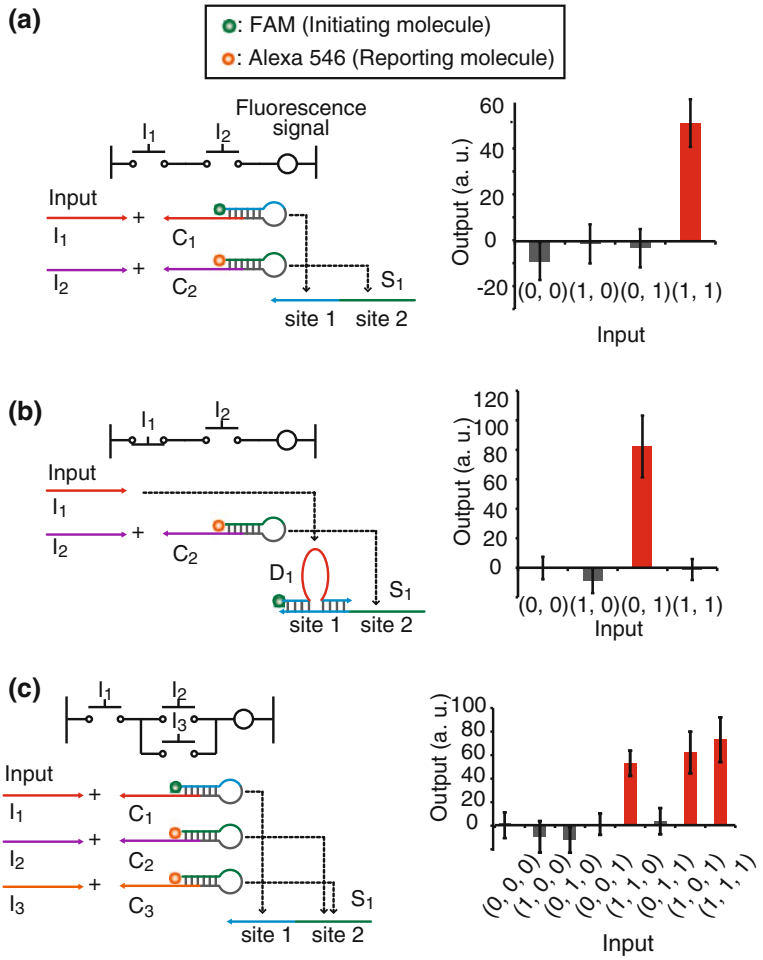


Fig. 4.15 Schemes and operation results for possible combinations of the inputs. **a** $I_1 \wedge I_2$, **b** $\neg I_1 \wedge I_2$, and **c** $I_1 \wedge (I_2 \vee I_3)$

are shown in Fig. 4.16. Two sites were formed on the scaffold. The individual sites can bind with the connecting DNAs for inputs I_1 and I_2. FAM and Alexa 546 (excitation/emission: 556/573 nm) were used as the FRET pair. When the two inputs I_1 and I_2 both exist, then the FRET path from FAM to Alexa 546 can be constructed on the scaffold. The concentrations of the input DNAs were 0 μM and 2 μM for the input values "0" and "1," respectively. The concentrations of the other components were 0.4 μM. The results for the possible combinations of the values of I_1 and I_2 are shown on the right side of Fig. 4.15a. The output signal is defined as the fluorescence intensity at the emission peak wavelength of the reporting dye when measured after adding the input DNAs, but minus the intensity measured before the addition.

S_1: 5' CATCGGGTGAGCGCTTCGGCAGAGCG 3'
S_2: 5' GCAACTATGAGCGCATCGGGTGAGCGCTTCGGCAGAGCG 3'
C_1: 5' FAM-CGCTCTGCCGAAGGCAGAGCGCCACTTACAA 3'
C_2: 5' Alexa546-CGCTCACCCGATGGGTGAGCGCTCAAGGATT 3'
C_3: 5' Alexa546-CGCTCACCCGATGGGTGAGCGAGGTAGTTGT 3'
C_4: 5' Alexa594-CGCTCATAGTTGCTATGAGCGAGGTAGTTGT 3'
D_1: 5' FAM-CGCTCTACCACTTACAAAGCCGAAG 3'
I_1: 5' TTGTAAGTGGCGCTC 3'
I_2: 5' AATCCTTGAGCGCTC 3'
I_3: 5' ACAACTACCTCGCTC 3'

Fig. 4.16 Sequences and modifications of the DNA strands used

The operational results show that a high fluorescence signal is obtained only when $(I_1, I_2) = (1, 1)$. This result indicates that the connecting DNAs work well and that the AND operation is executed successfully.

Figure 4.15b shows the scheme and the results of the logic function $\neg I_1 \wedge I_2$ used to demonstrate NOT operation. A disconnecting DNA which binds to site 1 was used for input I_1, and a connecting DNA which binds to site 2 was used for input I_2. The operation results show that a high output signal is obtained only for $(I_1, I_2) = (0, 1)$. The results demonstrate that operations including NOT are executed.

Figure 4.15c shows the scheme and the result of the logic function $I_1 \wedge (I_2 \vee I_3)$ used to demonstrate OR operation. For execution of the OR operation, the same fluorescent dye, Alexa 546, is assigned to inputs I_2 and I_3. The connecting DNAs for inputs I_2 and I_3 are different, but both DNAs can bind to site 2. As a result of this design, Alexa 546 is placed on site 2 when input I_2 or I_3 is "1". The result shows that high output signals are obtained for inputs of (1, 1, 0), (1, 0, 1), and (1, 1, 1). OR operation is therefore achievable based on the above design.

To demonstrate logic with more than 2 clauses, the logic function $I_1 \wedge I_2 \wedge I_3$ was executed. Figure 4.17 shows the scheme and the results. Three sites were formed on the DNA scaffold, and different connecting DNAs with three different fluorescent dyes, FAM, Alexa 546, and Alexa 594 (excitation/emission: 590/619 nm), were used. A high output signal was obtained only when $(I_1, I_2, I_3) = (1, 1, 1)$. This result demonstrates that the fluorescent dyes are arranged adequately according to the input and the logical operation with three clauses is executed successfully. The output signal is not low enough for inputs of (0, 1, 1), (1, 0, 1), and (1, 1, 0). Possible reasons for this are direct excitation of the second dye (Alexa 546) and FRET from the first dye (FAM) to the third (Alexa 594).

4.4.4 Analysis of Properties

Some properties of DNA scaffold logic were investigated by numerical analysis of the reactions. Let us consider the reactions between the input DNAs (I), the connecting

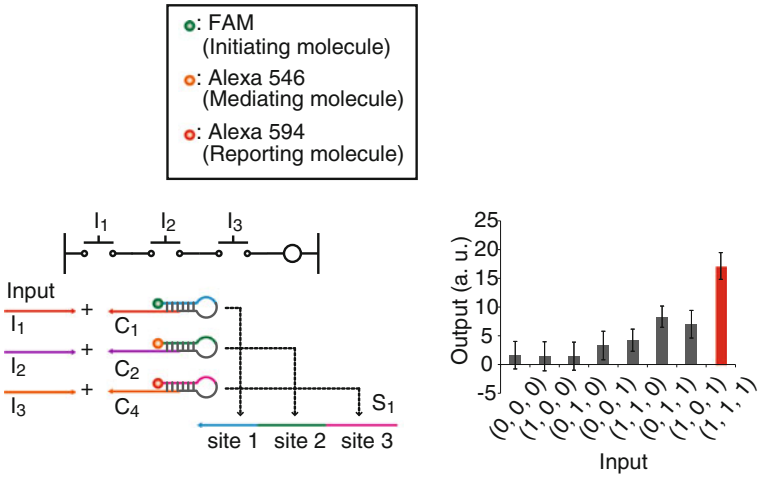


Fig. 4.17 The scheme and results for the logic function $I_1 \wedge I_2 \wedge I_3$

DNAs (C), and the DNA scaffolds(S). The reactions can be represented by chemical equations as shown in Fig. 4.18. Here, IC denotes a complex of DNAs I and C, while ICS denotes a complex of DNAs I, C, and S. k_1 is the reaction rate constant of the first reaction, and k_2 and k_3 are the reaction rate constants of the second reaction. The reaction rate equations are written as follows:

$$-\frac{d[I]_t}{dt} = k_1[I]_t[C]_t, \tag{4.1}$$

$$-\frac{d[C]_t}{dt} = k_1[I]_t[C]_t, \tag{4.2}$$

$$\frac{d[IC]_t}{dt} = k_1[I]_t[C]_t - k_2[IC]_t[S]_t + k_3[ICS]_t, \tag{4.3}$$

$$-\frac{d[S]_t}{dt} = k_2[IC]_t[S]_t - k_3[ICS]_t, \tag{4.4}$$

$$\frac{d[ICS]_t}{dt} = k_2[IC]_t[S]_t - k_3[ICS]_t, \tag{4.5}$$

$$[I]_0 = [I]_t + [IC]_t + [ICS]_t, \tag{4.6}$$

$$[C]_0 = [C]_t + [IC]_t + [ICS]_t, \tag{4.7}$$

Fig. 4.18 Chemical equations of the reactions between the input DNAs (I), the connecting DNAs (C), and the DNA scaffolds (S)

$$I + C \xrightarrow{k_1} IC$$

$$IC + S \underset{k_3}{\overset{k_2}{\rightleftarrows}} ICS$$

$$[S]_0 = [S]_t + [ICS]_t, \tag{4.8}$$

$$[IC]_0 = 0, \tag{4.9}$$

$$[ICS]_0 = 0. \tag{4.10}$$

Here $[\cdot]_t$ represents the concentration at time t, and $[\cdot]_0$ represents the initial concentration. From these equations, a single-variable differential equation on $[ICS]_t$ is derived as

$$\begin{aligned}\frac{[ICS]_t}{dt} &= \frac{k_2[I]_0[C]_0[S]_0(1-\exp\{([I]_0-[C]_0)k_1t\})}{[C]_0-[I]_0\exp\{([I]_0-[C]_0)k_1t\}} \\ &\quad -\left(k_2[S]_0+\frac{k_2[I]_0[C]_0(1-\exp\{([I]_0-[C]_0)k_1t\}}{[C]_0-[I]_0\exp\{([I]_0-[C]_0)k_1t\}}+k_3\right)[ICS]_t \\ &\quad +k_2[ICS]_t^2 \quad (\text{for } [I]_0 \neq [C]_0), \end{aligned} \tag{4.11}$$

$$\begin{aligned}\frac{d[ICS]_t}{dt} &= \frac{k_2[C]_0^2[S]_0k_1t}{[C]_0k_1t+1}-\left(k_2[S]_0+\frac{k_2[C]_02k_1t}{[C]_0k_1t+1}+k_3\right)[ICS]_t \\ &\quad +k_2[ICS]_t^2 \quad (\text{for } [I]_0 = [C]_0). \end{aligned} \tag{4.12}$$

This equation was solved numerically using the fourth-order Runge-Kutta method [28]. In the calculation, we simply assume that $k_1 = 3.4 \times 10^6\ \mathrm{M}^{-1}\mathrm{s}^{-1}$, $k_2 = 7.2 \times 10^4\ \mathrm{M}^{-1}\mathrm{s}^{-1}$, and $k_3 = 4.8 \times 10^{-4}\ \mathrm{s}^{-1}$. These were values estimated experimentally.

Figure 4.19 shows the dependence of $[ICS]_t$ on $[I]_0$, where $[C]_0 = 1.0\ \mu\mathrm{M}$, $[S]_0 = 1.0\ \mu\mathrm{M}$. $[ICS]_t$ indicates the concentration of the scaffold on which a fluorescent dye carried by DNA C is placed. At any time, $[ICS]_t$ is approximately proportional to $[I]_0$ when $[I]_0 < 1.0\,\mu\mathrm{M}$. Therefore, the ratio of the fluorescent dyes on the scaffold depends on the concentration of the input DNA. However, $[ICS]_t$ is almost constant when $[I]_0 \geq 1.0\ \mu\mathrm{M}$. This indicates the robustness of the behavior

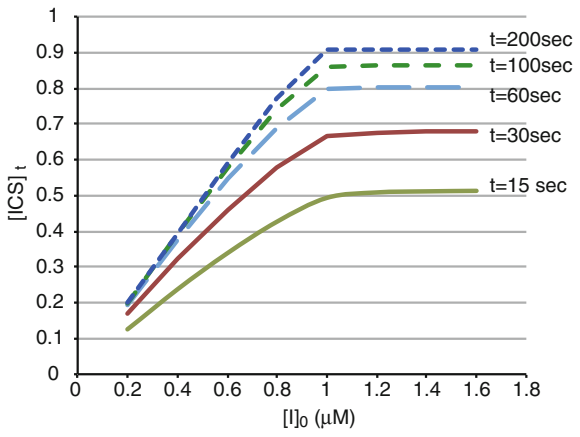


Fig. 4.19 Dependence of $[ICS]_t$ on $[I]_0$ at different times

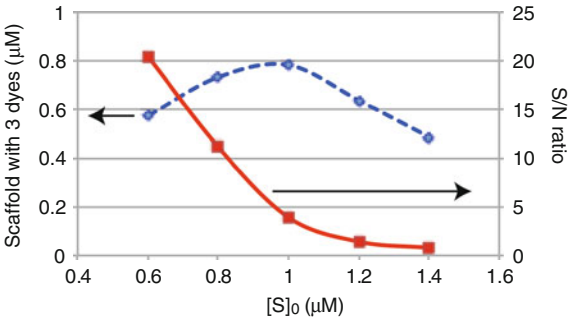


Fig. 4.20 Dependence of the amount of scaffolds with 3 fluorescence dyes (left axis) and the S/N ratio (right axis) on $[S]_0$

of $[ICS]_t$ for the variance of $[I]_0$, which is more than a value. The feature is useful for construction of digital circuits based on DNA scaffold logic.

Next, the signal to noise ratio (S/N) is estimated. When the DNA scaffold contains n sites, the ratio, $R_t(n, k)$, of the scaffold on which $k(\leq n)$ fluorescent dyes are placed is given by

$$R_t(n,k) = {}_nC_k \left(\frac{[IHS]_t}{[S]_0}\right)^k \left(1 - \frac{[IHS]_t}{[S]_0}\right)^{n-k}, \tag{4.13}$$

where ${}_nC_k$ is the k-combination from n elements. As described in Sect. 4.4.3 with regard to Fig. 4.17, when 2 out of the 3 dyes are placed on the scaffold, the reporting dye produces a measurable output signal as noise. We therefore define the S/N ratio as $R_\infty(n,n)/R_\infty(n,n-1)$. Figure 4.20 shows the dependence of the amount of scaffolds with n fluorescence dyes ($R_\infty(n,n) \times [S]_0$) and the S/N ratio on $[S]_0$ when $n = 3$, $[I]_0 = 1.0$ μM, and $[C]_0 = 1.0$ μM. The largest amount of scaffolds with 3 dyes is obtained around $[S]_0 = 1.0$ μM. Therefore, $[S]_0 = 1.0$ μM is a good choice under the conditions considered from the viewpoint of sensitivity for fluorescence detection. The S/N ratio, however, decreases with increasing $[S]_0$, which means that a smaller $[S]_0$ is better. These results indicate that the scaffold concentration should be determined carefully by considering the detector sensitivity and the acceptable S/N ratio.

4.5 Conclusions and Outlook

In this chapter, we presented the concept of the photonic DNA nano-processor and its implementation methods. This is a photonics-based approach to realize the molecular processing associated with physical processes such as sensing and actuation. In particular, we focused on the nano-processors that transform based on light-activatable sensing or that produce a FRET signal as a result of a logical operation based on molecular inputs. The basic concepts were demonstrated and some performance characteristics were then revealed by experimental and numerical analyses.

We then considered the future prospects of the photonic DNA nano-processor. We demonstrated the fundamental function of processing by the use of light-activatable nano-processors and using DNA scaffold logic. Functional extension is, however, required to realize valuable nanoscale information systems. For example, we are constructing a type of memory used in a FRET circuit [29]. The memory function is indispensable for an infinite automaton, which is more powerful than combinatorial logic. Communication among the nano-processors is also important to enhance the functionality. Along with molecular communication, broad-casting or broad-gathering using light will also be effective.

In Sect. 4.4.4, some characteristics were derived from analyses of the reactions. These findings support the issues for design of the photonic DNA nano-processors. Theoretical investigations, especially from the viewpoints of systems or information, will be useful for generation of design guidelines and for highlighting the technological issues in this field.

There are many potential applications of this technology. In particular, monitoring and regulating the activities of groups of cells is promising, because adaptive and parallel control depending on the status of the cells in question is possible based on the intermediate layer formed by the nano-processor between the nano-world and the macro-world. The photonic DNA nano-processor is expected to be a powerful tool in nanoscience fields including synthetic biology and drug delivery.

Acknowledgments This work was supported by Grants-in-Aid for Scientific Research (B) (No. 22300103) and for Scientific Research on Innovative Areas "Nanomedicine Molecular Science" (No. 2306) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

References

1. J.K. Willmann, N. Bruggen, L.M. Dinkelborg, S.S. Gambhir, Nat. Rev. Drug Discov. **7**, 591 (2008)
2. B. Huang, H. Babcock, X. Zhuang, Cell **143**, 1047 (2010)
3. N.C. Seeman, Annu. Rev. Biochem. **79**, 65 (2010)
4. E. Winfree, F. Liu, L.A. Wenzler, N.C. Seeman, Nature **394**, 539 (1998)
5. P.W.K. Rothemund, Nature **440**, 297 (2006)
6. E.S. Andersen, M. Dong, M.M. Nielsen, K. Jahn, R. Subramani, W. Mamdouh, M.M. Golas, B. Sander, H. Stark, C.L.P. Oliveira, J.S. Pedersen, V. Birkedal, F. Besenbacher, K.V. Gothelf, J Kjems. Nature **459**, 73 (2009)
7. S.H. Park, P. Yin, Y. Liu, J.H. Reif, T.H. LaBean, H. Yan, Nano Lett. **5**, 729 (2005)
8. W. Chongchitmate, C. Dwyer, A.R. Lebeck, IEEE Micro **30**, 110 (2010)
9. B. Yurke, A.J. Turberfield, A.P. Mills Jr, F.C. Simmel, J.L. Neumann, Nature **406**, 605 (2000)
10. W.B. Sherman, N.C. Seeman, Nano Lett. **4**, 1203 (2004)
11. J. Bath, A.J. Turberfield, Nature Nanotechnol. **2**, 275 (2007)
12. L.M. Adleman, Science **266**, 1021 (1994)
13. Y. Benenson, B. Gil, U. Ben-Dor, R. Adar, E. Shapiro, Nature **429**, 423 (2004)
14. M.N. Stojanovic, D. Stefanovic, Nat. Biotechnol. **21**, 1069 (2003)
15. G. Seelig, D. Soloveichik, D.Y. Zhang, E. Winfree, Science **314**, 1585 (2006)
16. L. Qian, E. Winfree, J. Bruck, Nature **475**, 368 (2011)

17. P.W.K. Rothemund, N. Papadakis, E. Winfree, PLoS Biol. **2**, 2041 (2004)
18. J. Elbaz, O. Lioubashevski, F. Wang, F. Remacle, R.D. Levine, I. Willner, Nature Nanotech. **5**, 417 (2010)
19. P. Cheben (ed.), *Proceedings of 2011 ICO International Conference on Information Photonics (on CD-ROM)*, Ottawa, 2011.
20. T. Nishimura, Y. Ogura, K. Yamada, H. Yamamoto, J. Tanida, Proc. SPIE **8102**, 810207 (2011)
21. T. Nishimura, Y. Ogura, K. Yamda, H. Yamamoto, J. Tanida, *in Abstracts for 8th Annual Conference on Foundations of Nanoscience* (Snowbird Cliff Lodge, Snowbird, Utah, April, 2011), pp. 181–182
22. H. Asanuma, X. Liang, H. Nishioka, D. Matsunaga, M. Liu, M. Komiyama, Nat. Protocols **2**, 203 (2007)
23. X. Liang, H. Nishioka, N. Takenaka, H. Asanuma, ChemBioChem **9**, 702 (2008)
24. Y. Ogura, T. Nishimura, J. Tanida, Appl. Phys. Express **2**, 025004 (2009)
25. C. Wang, Z. Zhu, Y. Song, H. Lin, C.J. Yang, W. Tan, Chem. Commun. **47**, 5708 (2011)
26. T. Nishimura, Y. Ogura, J. Tanida, Biomed. Opt. Express **3**, 920 (2012)
27. T. Nishimura, Y. Ogura, J. Tanida, Appl. Phys. Lett. **101**, 233703 (2012)
28. W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd ed. (Cambridge, England: Cambridge University Press, 1992), pp. 704–708.
29. T. Nishimura, Y. Ogura, J. Tanida, Appl. Phys. Express **6**, 015201 (2013)

Chapter 5
Boolean Logic Circuits on Nanowire Networks and Related Technologies

Seiya Kasai, Hong-Quan Zhao, Yuta Shiratori, Tamer Mohamed and Svetlana N. Yanushkevich

Abstract Implementation of graph-based logic circuits on semiconductor nanowire networks and related technologies are described. Boolean logic function is graphically represented utilizing a binary decision diagram (BDD), unlike the conventional logic circuit in which a logic gate has a specific function as a logic operator. Logical graph structure is topologically transferred to a semiconductor nanowire network structure. BDD 2-bit arithmetic logic unit (ALU) is demonstrated on a GaAs-based regular nanowire network with hexagonal topology. A reconfigurable BDD logic circuit based on Shannon's expansion of Boolean logic function and Fault-tolerant BDD on the basis of an error-correcting technique are also described.

5.1 Introduction

Unique network configuration of integrated nanomaterials and nanostructures has inspired us to use them for new information processing that the current Si CMOS logic LSIs do not cover. In the current LSI technology, devices and circuits are artificially designed in fine details and they are produced by perfectly controlled top-down nanotechnologies. They do not consist of nanostructures spontaneously produced by bottom-up nanotechnologies, even those with high regularity, high symmetry, and high density. However, recently, semiconductor nanowires are expected to be used as transistor channel materials in future LSIs beyond 10-nm technology [1, 2]. For moving our inspiration of the use of the nanomaterials and nanostructures to logic

S. Kasai (✉) · H.-Q. Zhao · Y. Shiratori
Graduate School of Information Science and Technology and Research Center for Integrated Quantum Electronics, Hokkaido University, N14, W9, Sapporo 060-0814, Japan
e-mail: kasai@rciqe.hokudai.ac.jp

T. Mohamed · S. N. Yanushkevich
Department of Electrical and Computer Engineering, University of Calgary, Calgary, Alberta, T2N 1N4, Canada

M. Naruse (ed.), *Nanophotonic Information Physics*,
Nano-Optics and Nanophotonics, DOI: 10.1007/978-3-642-40224-1_5,

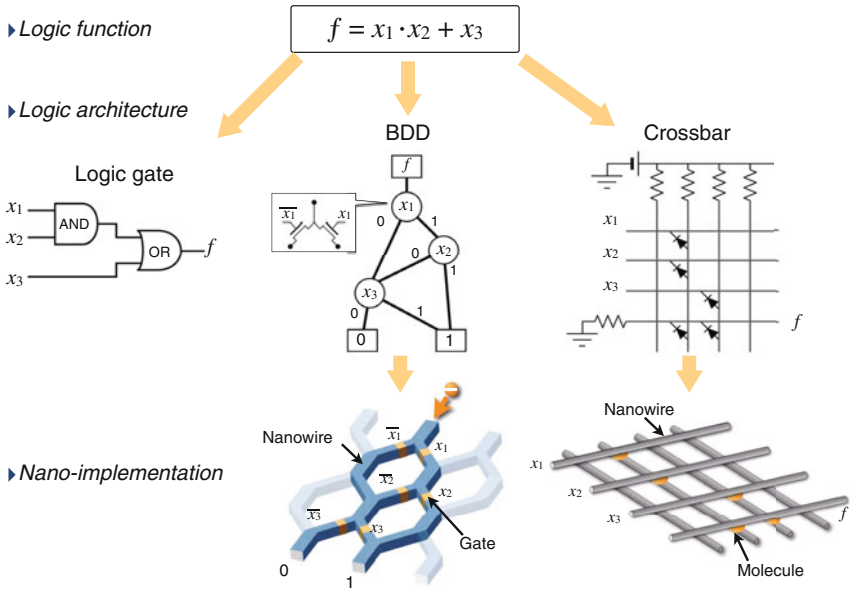


Fig. 5.1 Logic circuit for nanomaterials and nanostructure networks

circuits, the most important issue is to find out a logic architecture bridging between the physical network structure and the logic function described by Boolean algebra. In this section, we describe important examples, binary-decision diagram (BDD) and crossbar. Both architectures are suitable for a nanowire network structure.

The BDD is a representation scheme of a logic function using a directed graph as shown in Fig. 5.1 [3, 4]. On the basis of the Shannon expansion, any Boolean function can be represented with a binary decision tree [4]. Each node has an input branch and two exit branches and switches the path of a messenger entering from the entry branch according to an input value, x. The reduced graph utilizing reduction techniques is usually called as BDD. The BDD is known to realize compact representation of a data structure [4]. The function is evaluated by propagating a messenger from a root to terminals (leaf of the graph tree) and identifying which terminals the messenger reaches, labeled "0" or "1".

The crossbar architecture is a two dimensional array of intersecting sets of orthogonal wires as shown in Fig. 5.1 [5]. A Boolean function is represented with connecting or disconnecting wire-crossing nodes. The crossbar is regarded as a memory array. A matrix of junctions can represent the combination of AND gates in a simple way. On the other hand, to implement OR gates, peripheral components are necessary. To represent the complete set of the Boolean algebra, complementary input signals are necessary.

Table 5.1 summarizes the main features of the BDD and the crossbar. The advantage of the BDD is that the logic function is graphical and visible. The graph can

Table 5.1 Logic architectures for network structure and their features

Logic architecture	Physical structure	Logic representation	Function of junction node	Advantage	Disadvantage
BDD	Binary tree and its topological structures	Graph	Path switch	Compact flexibility in structure selection	Wirings for input accessing nodes
Crossbar	Grid	Matrix	Diode	Simple	Peripheral components are necessary reliability/defect

be topologically implemented on the various nanowire networks. This makes it applicable to various network structure produced by the bottom-up nanotechnology. Either circuits have both advantages and disadvantages. In future, we will select suitable one in accordance with available material, process, size, reliability, cost, and so on of these architectures.

Nano-scale implementations of the BDD and the crossbar are also shown in Fig. 5.1. In a nano-scale crossbar, the physical circuit is implemented by the nanowire or carbon nanotube array and the conduction of each junction is controlled by inserting suitable molecules [5]. Realization of programmable junction nodes is an important factor. It is difficult to physically access and control the connection at individual nodes. Then, electrically controllable two-terminal nonvolatile memory is inserted in every node. An addressed node is switched by applying suitable voltage/current on the junction. This concept is similar to the programmable logic array (PLA). Certain molecules have bistable states and can work as a memory device [5]. Atomic switch has been also developed for such purpose [6]. It is found that nano-scale junction of metals itself works as a non-volatile memory [7]. On the other hand, reliability related to the defect of the junction is a problem. Fault tolerance such as the majority logic architecture using several redundant circuits has been investigated [8].

For nano-scale implementation of the BDD, a nanowire-network based logic circuit was proposed and has been developed [9–12]. In this circuit, the graph is directly implemented on a regular nanowire network having hexagonal topology [9, 10]. This architecture matches regular nanowire network structures. It simplifies circuit design, layout, device structure, and the fabrication process. Feasibility of this implementation scheme has been clarified by successful demonstration of small logic functions [10, 13], a 2-bit full adder [14], and implementation of a signal processing system by circuit simulation [11]. It is also noted that the BDD enables us to use quantum nanodevices, such as rapid single flux quantum [15, 16], quantum wire [17], and single electron devices [10, 18], even though these devices have difficulties in implementing the conventional logic gate architecture due to small transfer gain and

small current drivability. Topological transformation of the graph provides us good opportunity to use various bottom-up nanowire network structures [4]. Recently, the addition of reconfigurable capability has been investigated [19]. Comparing with the crossbar circuit, the disadvantage of the BDD circuit is many wirings accessing to each node for input signals. By suitable design of the network, the wiring layout can be simplified. Details are described later.

5.2 BDD Logic Circuit on Nanowire Network

5.2.1 Basic Concept and Nanoscale Implementation

As compared with the conventional logic gate architecture, the BDD is a more appropriate architecture for representing a Boolean logic function on a nanowire network, since it is a graphical way to represent a logic function and has a simple structure [4, 20]. The BDD can be mapped directly to the electronic circuit by exchanging each switching node with a multiplexer and so on. A BDD is a rooted directed graph, derived from a binary decision tree as shown in Fig. 5.2, representing a logic function via Shannon expansion,

$$f(x_0, x_1, \ldots x_i, \ldots) = \bar{x}_i f(x_0, x_1, \ldots, 0, \ldots) + x_i f(x_0, x_1, \ldots, 1, \ldots) \quad (5.1)$$

where $f(x_0, x_1, \ldots, 0, \ldots)$ and $f(x_0, x_1, \ldots, 1, \ldots)$ correspond to the exit branches labeled "0" and "1" of the node controlled by x_i. Let consider a Boolean function $f = x_1x_2+x_3$ for example. Its truth table is shown in Fig. 5.2a. A binary decision tree (BDT) representing f is shown in Fig. 5.2b. This tree has a data structure completely matching the truth table of f. The BDT can represent the function compactly. The

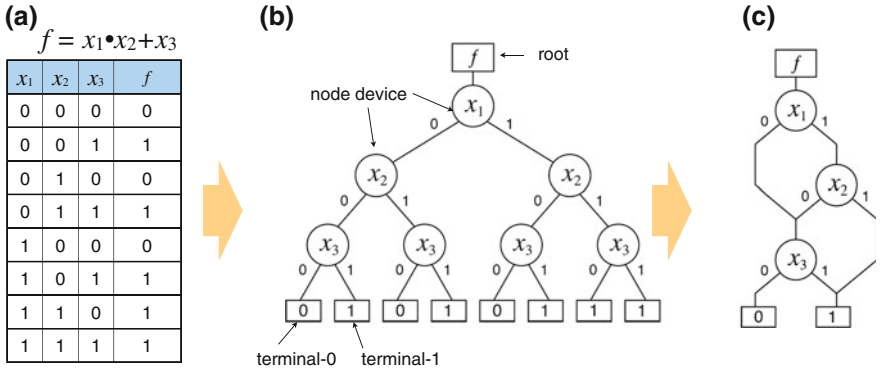


Fig. 5.2 Concept of the BDD. **a** Truth table of $f = x_1x_2 + x_3$, **b** binary decision tree of f, and **c** binary decision diagram with hexagonal topology

tree consists of a root, terminals-0 and 1, and a number of path-switching nodes labeled x_i. Each node has two exit branches labeled "0" and "1". The logic value of the function can be determined for a given input variable assignment by following a path down the tree from the root to a terminal. In each node, one of exit branches is selected according to the input binary variable, x_i. The cost of implementing a path-switch circuit is quite low in pass-gate CMOS design where the node consists of a pair of pass transistors [21]. Note that the BDD can express exclusive OR (ExOR, XOR), which is frequently used in basic arithmetic operations, more compactly than the CMOS logic gate.

Nano-scale physical implementation of the BDD logic circuit is schematically shown in Fig. 5.1d. We usually transform the binary decision tree into a binary decision diagram with hexagonal topology, as shown in Fig. 5.2c, by several reducing and ordering techniques [4] together with topology control. Closely packed integration of node devices having a threefold symmetric configuration naturally results in the hexagonal topology of the graph network. Any combinational logic function can be represented in this way. The logical structure in Fig. 5.2c is directly transferred to a semiconductor-based nanowire network having the same topology. As shown in Fig. 5.3, the BDD circuit can be implemented on the various physical structures by topological control of the logical structure [23–27]. To evaluate the function, the messenger electrons are sent from the root to the terminals. The logic value is determined by observing whether the electrons reach the terminal-1. If the electrons are detected in terminal-1, the logic is true. If not, it is false. This can be easily performed by standard electrical current measurement. The logic function can be evaluated only using terminal-1, so we usually omit terminal-0. The path switch function of the node is implemented by attaching a nanometer-size gate to each exit nanowire branch, as shown in Fig. 5.4a, which controls nanowire conduction by the field effect. The device is usually called as node device. The path is switched by giving complementary gate voltages, V_{x1} and $-V_{x1}$, simultaneously to the two gates in a complementary fashion. The node devices with quantum wire transistors (Fig. 5.4b) or single-electron transistors (Fig. 5.4c) are also available which potentially operate with a small number of electrons, resulting in ultra-low power consumption [12]. We have also developed a path switch device operating without complementary input singal [28].

There are mainly three reasons that this circuit approach is applicable to regular nanowire networks: the regular graph structure without additional branch crossover, passive operation, and the simplified structure and fabrication process. If a large-scale high-density regular network with hexagonal topology can be formed, the circuit area can be simply reduced. Any combinational logic circuit can be fabricated only by removing unnecessary branches and by attaching gates. This provides an opportunity to use regular network structures spontaneously formed by bottom-up nanotechnologies.

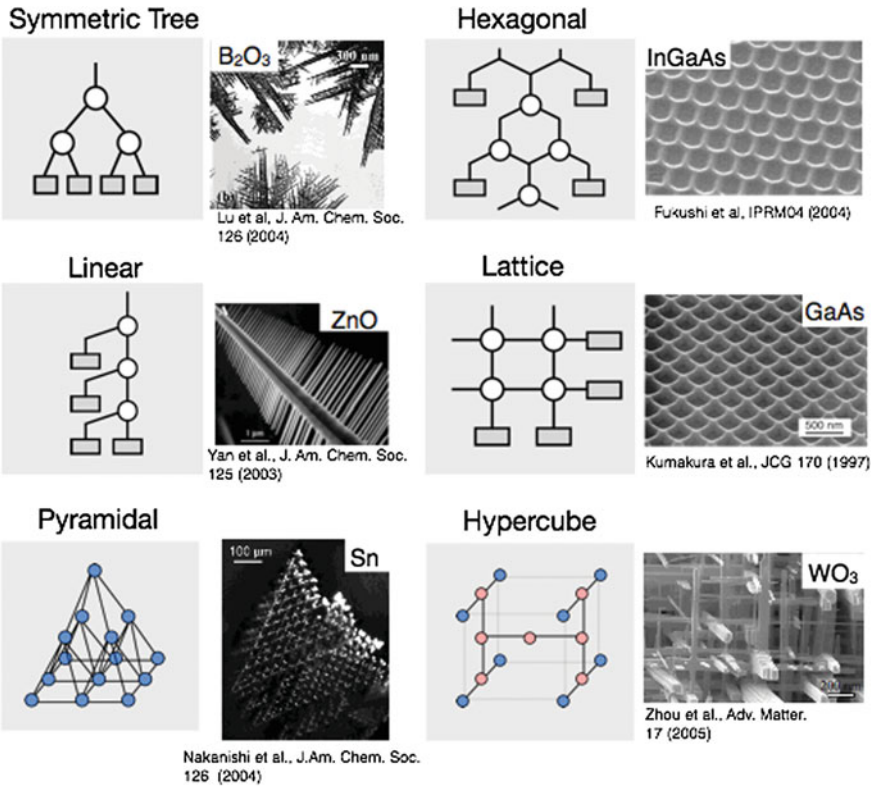


Fig. 5.3 Logical structures of BDD with topologically difference and corresponding physical nanowire network structures [23–27]. ([22], [24], and [26] with permission from American Chemical Society, [23] with permission from Springer, [25] with permission from Elsevier, [27] with permission from John Wiley and Sons)

5.2.2 Circuit Design

In this section, the circuit design of the nanowire-network BDD is explained taking an arithmetic logic unit (ALU), which is a building block of a central processing unit, as an example. We show that the ALU can be designed by the nanowire-network BDD with a simple and regular logical structure.

The ALU integrates a set of subsystems. A BDD diagram of a 4-instruction 2-bit ALU is shown in Fig. 5.5a. The circuit design and structure are simplified by bit-slice design [29] and introducing some redundancy. The design procedure is as follows. First, an instruction set is designed. Next, BDD subsystems (i.e., subgraphs) implementing each instruction are designed. A merging rule is applied to reduce the isomorphic nodes, retaining no additional nanowire crossover. The order of variables is arranged to be the same in all subgraphs. Then node devices receiving the same logic input are aligned in the same horizontal line. After that, the subgraphs are

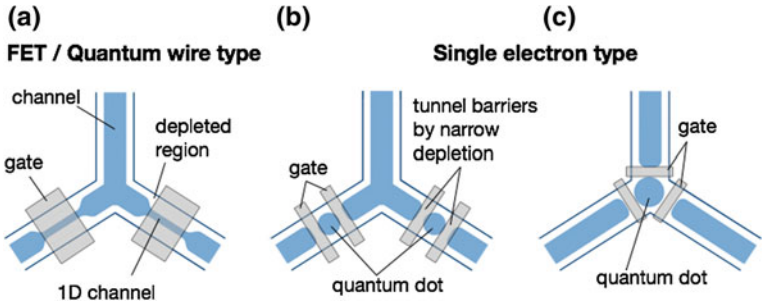


Fig. 5.4 Electron BDD node devices. **a** FET/Quantum wire type, **b** single-electron type with two quantum dots and **c** single quantum dot

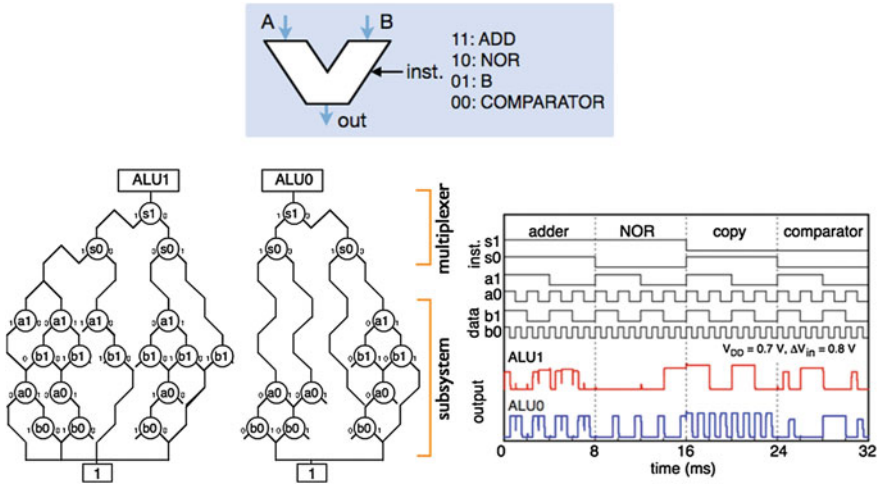


Fig. 5.5 BDD-based 4-instruction 2-bit arithmetic logic unit (ALU). **a** Circuit diagram with hexagonal topology and **b** simulated input-output waveforms obtained from a conventional circuit simulator

integrated using a multiplexer tree of each bit of the output. Finally, the lateral order of the subgraphs is arranged to decrease the total area. The present design approach keeps the logic function for each instruction clearly visible and makes the design and node device layout simple. The shared BDD technique [30] will also reduce the graph size and area. Various techniques and tools for synthesizing BDD and reduction of logical graph size are available [4].

The ALU shown in Fig. 5.5a implements four instructions: Adder, NOR, Copy, and Comparator. A 4-input multiplexer combined them for each bit of the output labeled ALU1 and ALU0. Smaller device counts than that of CMOS logic gate architecture is expected in many cases because the BDD represents logic functions compactly [31]. The node device count in the present ALU is 32 with 45 field-effect gates, although the transistor count in a straightforward CMOS design with

the same composition is 136. A small device count helps to reduce the total area and to compensate introduced redundancies for simplifying the design. Typical activity factors α of switching, switching probability per clock, in a static CMOS logic gate and the BDD circuit are 0.1 [32] and 0.2 [14], respectively. The total active power consumption of the system, P_A, is estimated by the next formula,

$$P_{AC} = \alpha \cdot N_G \cdot CV^2 f_{clock} \tag{5.2}$$

where N_G is the device count (the number of field effect gates), C is the switching capacitance, V is the switching voltage, and f_{clock} is the clock frequency. Assuming the same elemental device performance (CV^2) at the same clock frequency, the BDD is expected to give smaller power consumption, since $\alpha N_G = 9$ of the BDD is smaller than that of CMOS, $\alpha N_G = 13.6$ in the case of the present ALU design.

The operation of the present BDD ALU was examined by circuit simulation using a conventional circuit simulator. Each node device was represented using two FETs. The device parameters for simulation were extracted from DC and RF measurements [33]. Obtained waveforms are shown in Fig. 5.5b. Correct outputs were obtained in the designed circuit. In this simulation, the circuit was operated asynchronously. Synchronous circuit design is important for current large-scale integrated logic circuits. Timing control by conventional design inserting flip-flop (FF) circuits is possible. We have developed the FF which can be implemented on the same nanowire network [34, 35]. Another possible design is to use clocked electron transfer utilizing a charge-coupled device (CCD). It only needs additional gates operating as a turnstile synchronized with the clock. We have already demonstrated CCD operation in the multiple-gate nanowire with a few hundred electrons [36]. An example of the fabricated CCD on the GaAs-based nanowire and the charge transfer characteristics in Fig. 5.6. This design is simple and seems suitable for the present BDD. Single- or a-few-electron transfer has been demonstrated experimentally [10, 17, 37] and this type of the synchronous single electron BDD circuit was successfully implemented using the circuit simulation [38].

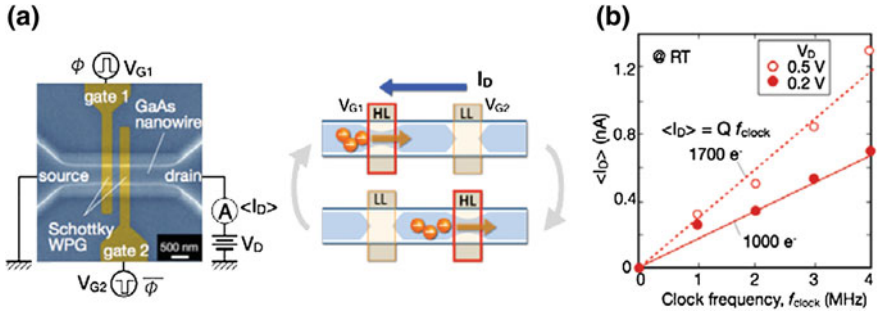


Fig. 5.6 **a** Two-gate CCD on a GaAs-based nanowire with its operation principle and **b** measured output DC current as a function of clock frequency

5.2.3 Fabrication

For nano-scale implementation of the present BDD circuit, a GaAs-based nanowire network is utilized as a host network structure. The present circuit can be fabricated using a standard III-V semiconductor device fabrication process and no special technique is necessary. As a host structure, a regular nanowire network with hexagonal topology is formed on a conventional AlGaAs/GaAs modulation-doped heterostructure on a (001) GaAs substrate by electron beam (EB) lithography and wet chemical etching. The $\langle \bar{1}10 \rangle$, $\langle 100 \rangle$, and $\langle 0\bar{1}0 \rangle$directions are chosen for nanowire branches, because crystallographic low-index direction realizes a straight structure with smooth sidewalls in atomic scale. The smooth surface is important for formation of the Schottky gate with ideal gate control characteristic. The structure design with low-index facets also greatly helps to form smooth and uniform networks reproducibly in a wide area. Etching depth is deep enough to reach the two-dimensional electron gas (2DEG) channel, which is necessary for electrical isolation of each device. The fabricated nanowire width was typically a few hundred nm. Ohmic contacts are formed for the roots and terminals. Schottky wrap gate (WPG) is formed on the nanowire for path switching, in which a metal gate wraps around the nanowire. Typical gate length is 400 nm. Interconnect metal lines with 100 nm-width or less are formed at the same level as WPGs without insulators. They do not work as gates because the threshold voltage of the narrow metal lines is very negative due to the short channel effect [38].

5.2.4 Device Characteristics

An example of the fabricated node device is shown in Fig. 5.7a. Scanning electron microscope (SEM) image shows that a smooth nanowire junction is formed by wet chemical etching. The WPG-controlled nanowire is preferable not only in a simple fabrication process but also in low active switching power. The WPGs formed on the two exit branches exhibited $I - V$ characteristics similar to the conventional FET at room temperature, as shown in Fig. 5.7b. Their characteristics are similar to each other. The maximum transconductance, g_m, is 161 mS/mm at source-drain voltage V_{DS} of 0.2 V in devices with nanowire width $W = 570$ nm and gate length $L_G = 550$ nm. The device structure and process is designed to operate the device in enhancement mode. A histogram of measured threshold voltage, V_{th}, for 16 devices having nominally same device dimensions on 8 different wafers is shown in Fig. 5.7c. Mean value of V_{th} is 0.23 V and the standard deviation of the threshold voltage σV_{th} is 0.024 V. The circuit can operate correctly by setting sufficient large input voltage swing to compensate the V_{th} variation. Measured subthreshold slope of the WPG-controlled nanowire is shown in Fig. 5.8 as a function of temperature. The slope follows an ideal curve given by $kT \ln(10)/e$ in a wide temperature range. Slight deviation in low temperatures is probably due to the tunneling effect around the top of the

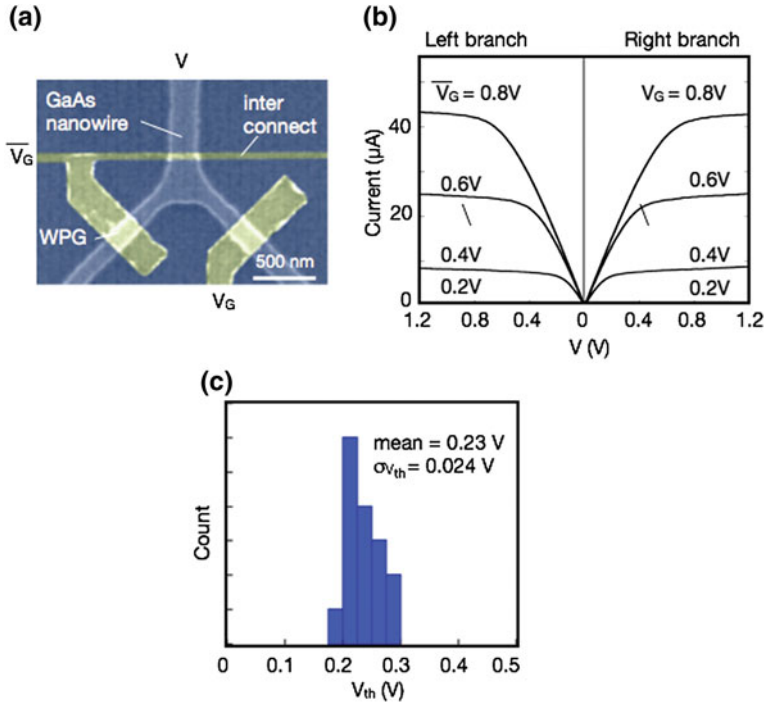


Fig. 5.7 **a** Fabricated node device, **b** measured I-V characteristics in the left and right branches controlled by WPG, **c** histogram of the threshold voltages from the 16 fabricated devices

potential barrier [39]. Simulated curves for the WPG and planar gate structures using numerical three-dimensional (3D) potential simulation are also plotted in Fig. 5.8 by broken lines. The theory reproduces the experimental data reasonably well. The simulation clarified that the nearly ideal subthreshold characteristic of the WPG on the nanowire is owing to the tight potential control resulted from the three-dimensional gate configuration.

The WPG can realize a one-dimensional channel by squeezing the nanowire electrostatically with suitable gate bias [17, 38]. Figure 5.9 shows the typical transfer curve at low temperature. Clear conductance quantization is observed, confirming the one-dimensional quantum transport. Conductance quantization was found to appear even at 100 K if the geometrical nanowire width was decreased [40]. Over 80 % of fabricated devices successfully showed conductance quantization at 30 K when $W \leq 200$ nm [40]. The WPG device potentially operates with a possibly ultra-small input voltage swing determined by the abrupt quantized conductance, also resulting in an ultra-low power consumption of the BDD circuit.

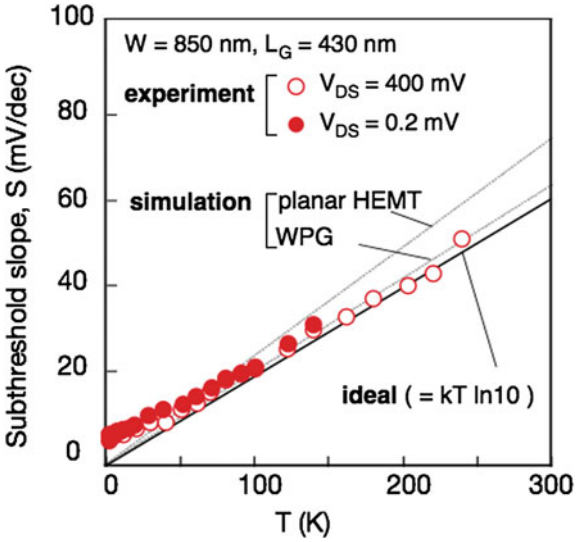


Fig. 5.8 Threshold slope in a WPG-controlled nanowire FET. Theoretical curves for the nanowire FET and a planar FET from 3D potential simulation are also plotted by dotted lines

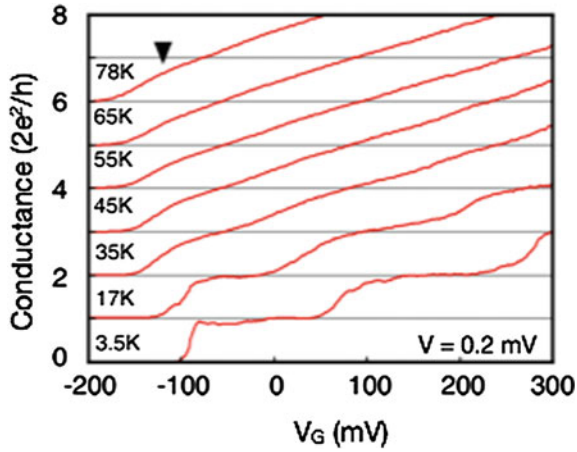


Fig. 5.9 Conductance quantization in WPG-controlled nanowire device at low temperatures

5.2.5 Circuit operation

A SEM image of a fabricated ALU is shown in Fig. 5.10a. The circuit integrates 32 node devices using 3 M nodes/cm^2 fabrication process. The size of each hexagon is $6 \times 6\,\mu m^2$. The total circuit area is $70 \times 45\,\mu m^2$. L_G and W are 550 and 570 nm, respectively. The fabrication process of this circuit is completely the same as that for discrete node devices. The circuit area of the BDD ALU depends on the density of the nanowire network. A higher density fabrication process of 45 M nodes/cm^2 has already been developed [41]. This reduces the circuit area less than 10 % of the present circuit. In case of the present material and structure, a critical factor in the

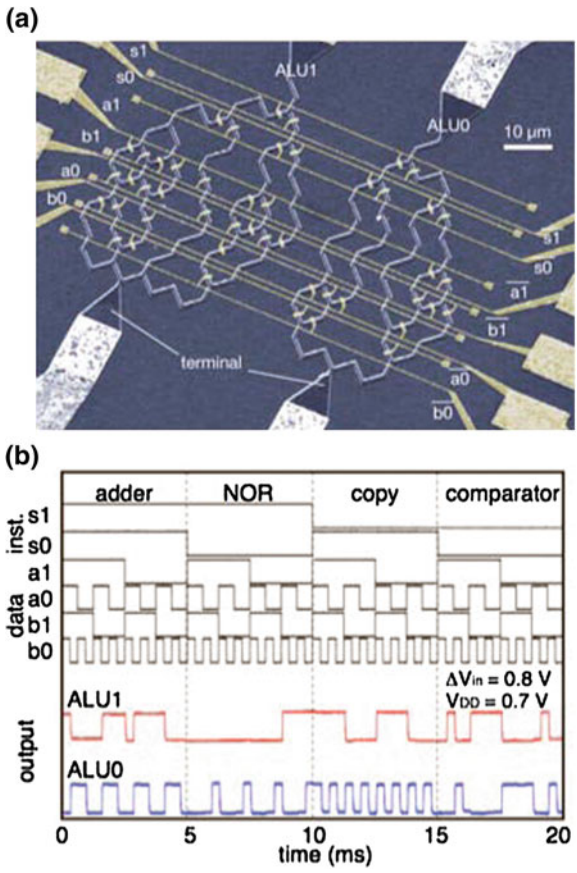


Fig. 5.10 **a** SEM image of the fabricated 2-bit ALU and **b** measured input-output waveforms

scale down is the nanowire width, because the side depletion of the nanowire limits the effective channel width around 40 nm [40]. Then a possible node density is 1G nodes/cm^2 where the circuit area can be 1/400 of that in Fig. 5.10a. Narrow and conductive semiconductor nanowires and networks are expected to be produced by other sophisticated nanotechnology [42–44].

The fabricated circuit operation was characterized at room temperature. Figure 5.10b shows measured input-output waveforms. Supply voltage, V_{DD}, of -0.7 V was applied to the roots to send electrons from the roots to the terminals. Output currents in terminals were converted to voltages through 1 MΩ resistors and they were monitored using a conventional oscilloscope. A DC offset voltage, V_{offset}, of -0.2 V was applied to all the WPGs except the first level of the nodes, s_1, which were biased at different voltage to make the height of the output waves uniform. An input voltage swing, ΔV_{in}, was 0.8 V. The output waveforms were found to reproduce the simulated result in Fig. 5.5b. Thus the fabricated ALU was found to operate

correctly. The obtained result clarifies the feasibility of the nanowire network BDD circuit for achieving a logic circuit at ALU level functionality.

5.2.6 Performance

In this section, we discuss about the circuit performance to see whether the nanowire-network BDD circuit offers any merit. We estimate speed and power consumption on the basis of the information from the fabricated circuit. Execution time of the circuit for asynchronous operation is estimated from a simple *RC* chain. In a general case, *C* is determined by WPG gate capacitance, C_{G}. The resistance *R* can be evaluated from the resistance per unit length of the nanowire, R_{NW}, and the physical path length from a root to a terminal. Then the execution time of the present ALU, τ_{ALU}, is estimated by

$$\tau_{ALU} = 2R_{NW}L_{node}C_G\left\{\left[\log_2 N_S\right] + N_B\right\}, \tag{5.3}$$

where N_{S} is the number of instructions, N_{B} is the number of bits of data, and the brackets denote a floor function. The term in the braces denotes the number of nodes between a root to a terminal. *RC* delay in each node is $R_{\mathrm{NW}}L_{\mathrm{node}}C_{\mathrm{G}}$, where L_{node} is the physical nanowire length in a node. The fabricated ALU had $N_{\mathrm{S}} = 4$, $N_{\mathrm{B}} = 2$, and $L_{\mathrm{node}} = 6\,\mu\mathrm{m}$. R_{NW} was $3\,\mathrm{k}\Omega/\mu\mathrm{m}$ [34]. C_{G} of a 550 nm-gate WPG was estimated to be 3.0 fF from the result of RF characterization [33]. Using the parameter values above, $\tau_{\mathrm{ALU}} = 0.43\,\mathrm{ns}$ was obtained, corresponding to the intrinsic operation frequency of 2.3 GHz. The execution time will be decreased as the density of the nanowire network increases, mainly because L_{node} decreases. Even narrowing the nanowire, product $R_{\mathrm{NW}}C_{\mathrm{G}}$ is unchanged. R_{NW} increases as *W* decreases, but C_{G} decreases and cancels out.

The power-delay product (PDP) of the ALU in terms of active switching is estimated by

$$PDP = \frac{N}{N_S}C_G\Delta V_G^2. \tag{5.4}$$

where *N* is the node device count. N/N_{S} is an average number of node devices per instruction. The possible PDP of the present ALU with $N = 32$ and $\Delta V_{\mathrm{in}} = 0.4\,\mathrm{V}$ is estimated to be $3.8 \times 10^{-15}\,\mathrm{J}$. This value is comparable to the PDP of the reported full adder using an ultra-low power subthreshold CMOS circuit with 350 nm design rule [45], indicating the low-power consumption capability of the BDD ALU. Assuming the simple scaling rule of the performance, the power consumption of the BDD ALU is expected to be 64 % of the subtreshold CMOS ALU with the same design rule. As shown in Fig. 5.9, the WPG-controlled nanowire operated as a quantum wire transistor at low temperature. The present BDD circuit is also expected to operate with the ultra-low power consumption in such quantum transport domain. In this case, the input voltage swing can be decreased because the device switches a smaller number of electrons using the abrupt quantized conductance step, offering small

voltage operation. Possible PDP in the WPG-controlled nanowire FET operating as a quantum wire transistor at 30 K was 3×10^{-20} J for the device with $L_G = 300$ nm [40]. This results in the PDP of 2.4×10^{-19} J for the ALU.

The impact of V_{th} variation on the circuit operation is a very important issue in the nano-scale logic circuit, even for the conventional CMOS circuit, because it limits the reduction of voltage and power consumption. However, the variation becomes remarkable as the device size is reduced. The effect of the variation on the present circuit is investigated through characterizing the input-voltage-swing dependence of the output voltage. The BDD circuit in the Fig. 5.1 can operate accepting V_{th} variation, if the input voltage swing is large enough to cancel the variation. Measured output voltage amplitude, ΔV_{out}, is shown in Fig. 5.11. In this measurement ΔV_{in} and V_{offset} was the same for all inputs. V_{DD} and V_{offset} were fixed at 0.7 and -0.2 V, respectively. Input voltage at which WPG switched on, V_{on}, was evaluated by $\Delta V_{in} - V_{offset}$. The circuit could operate when ΔV_{in} was as small as 0.4 V. From the analysis on the behavior shown in Fig. 5.11, we found that the WPG having the smallest V_{th} turned on at $V_{on} = 0.2$ V. This value was reasonably consistent with the average V_{th} of 0.23 V in Fig. 5.7c. The saturation of the output around 0.6 V was caused by the inserted resistor of 1 MΩ. In the fabricated circuit, the V_{th} variation was found to be small and large input voltage swing was not necessary.

Toward a practical system utilizing the BDD circuits, sequential circuits such as flip-flops should be also implemented on the same host network. A problem is that the sequential circuits are severely influenced by the V_{th} variation at low supply voltage. As a possible approach, we have investigated three-branch nanowire-junction (TBJ) device [36, 46–49]. The device has a completely same structure with the node of the network used for the present BDD. It shows a unique nonlinear voltage transfer characteristic [50] and it can work as AND gate by itself [27, 48]. Particularly such useful function can be realized without any field effect gate. By integrating NOT gate with two nanowire FETs, NAND gate with voltage transfer gain larger than unity can be realized [48]. Recently we demonstrated a novel flip-flop circuit integrating

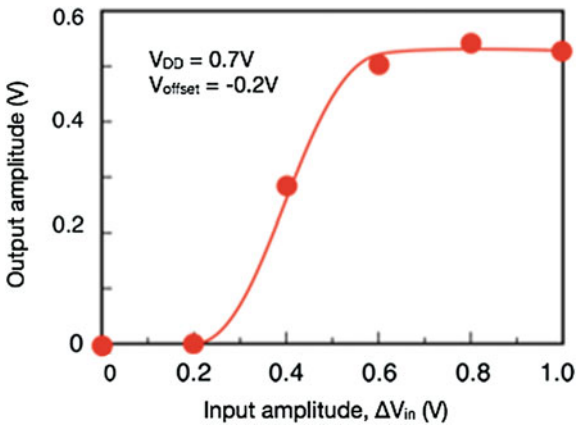


Fig. 5.11 Measured amplitude of output voltage as a function of input amplitude

TBJs [36]. This reduces the risk of out of operation due to the threshold voltage variation, because the number of the field-effect gates is decreased. In addition, for very weak signal readout in fluctuation, a bio-inspired approach that utilizes stochastic resonance (SR) in nanowire devices has been started to investigate [51, 52]. The SR is a phenomenon in which the response of a system is enhanced by noise and thermal fluctuation.

The present BDD circuit is potentially implemented by other devices, such as nanophotonic devices, and molecular devices as well as quantum devices. Major factors preventing their application to the conventional integrated circuits are low transfer gain and large variation. Since the present circuit operates with passive switching, the first problem is party solved, whereas level matching is necessary between subsystems. This has been confirmed by successful demonstration of BDD circuits integrating single electron devices [10] and quantum wire devices [17]. The variation can be also compensated in the present circuit by large input amplitude. From Fig. 5.11 and discussion of the data, the circuit was found to operate correctly even though each device did not turn on sufficiently and there were only small signal power.

5.3 Reconfigurable BDD Circuit

5.3.1 Concept

In this section, we describe our nanowire network reconfigurable BDD circuit architecture. One of the problems in the application of nanowires and carbon nanotubes produced by the bottom up nanotechnology is position control [1]. The nanowire-network BDD is favorable for such materials rather than the conventional CMOS logic circuit, since it consists of a regular array of elemental devices whereas the CMOS has a free-layout transistor array [53]. An approach to further compromising on the bottom-up type material, is implementation of the circuit reconfiguration capability, where logic functions are changed or programmed without rearranging the physical layout. Even though the produced network configuration is different from the desired one, we can adjust the function by programming. The reconfigurable BDD circuit architecture has universality even with its simple structure, because it is based on Shannon's expansion of a Boolean function and its graphical representation [19]. A reconfigurable circuit of two-variable Boolean logics is demonstrated by fabrication of the circuit on a GaAs etched nanowire network.

The basic concept of the reconfigurable BDD circuit is shown in Fig. 5.12. The previous nanowire-network BDD circuit having a graph of a specific logic function is fabricated by physically disconnecting unnecessary branches on a regular nanowire network [9–12]. The idea of the reconfiguration is to make disconnection of the nanowire branches "electrically" in accordance with a program. In order to secure the universality of this circuit concept, we consider a Boolean function expanded by

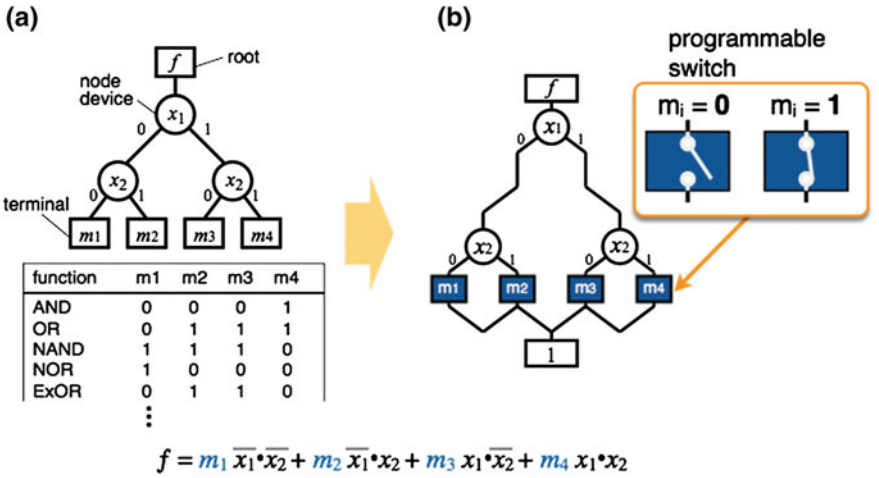


Fig. 5.12 Basic concept of the reconfigurable BDD. **a** Binary decision tree for the universal expression of two-variable Boolean functions and **b** its BDD representation

Shannon's expansion. In case of a two-variable Boolean logic function for example, its expanded form is given by,

$$f = m_1 \cdot \bar{x}_1 \cdot \bar{x}_2 + m_2 \cdot \bar{x}_1 \cdot x_2 + m_3 \cdot x_1 \cdot \bar{x}_2 + m_4 \cdot x_1 \cdot x_2 \tag{5.5}$$

where x_1 and x_2 are variables, and m_j (j = 1, 2, 3, 4) is a coefficient that takes a value of 0 or 1. A complete set of Boolean algebras is expressed with all combinations of m_j: there are 2^4 combinations for the two-input case. Major logic functions and corresponding combinations of m_j are summarized in Fig. 5.12a. The expanded function can be represented graphically using a binary decision tree as shown in Fig. 5.12a. The graph consists of a root, nodes, and terminals. Each node represents a variable x_i(i = 1, 2). Each terminal corresponds to m_j denoted by 0 or 1, which mean false and true of the logic function, respectively. All of 2-variable logic functions ($= 2^4$) can be represented by this graph. The physical logic circuit is implemented by topologically transferring the logical graph structure to a physical nanowire network as shown in Fig. 5.12b.

The reconfigurable BDD circuit integrates programmable switches together with a root and node devices. Each programmable switch is physically connected to the terminal-1. When $m_j = 1$, the switch is turned on and the nanowire is electrically connected to the terminal-1. The logic value can be evaluated by measuring the current between the root and the terminal-1. The current does not flow, then the logic is false. The logic can be evaluated by watching the terminal-1 whether the current flows or not. Thus the terminal-0 can be omitted. By setting m_j, every two-variable logic function can be implemented without changing the physical circuit configuration. If the switch can be programmed in a short time, dynamic reconfiguration

becomes possible. The numbers of node devices and programmable switches for n-input variable functions are $2n - 1$ and $2n$, respectively. When the number of implemented functions is reduced, size of the circuit can be made smaller by using the reduced-order BDD technique [4]. The proposed circuit architecture is simpler and more compact than the previously reported one in which every node device is programmable [54].

5.3.2 Fabrication and Characterization

Figure 5.13a shows schematics of the programmable switch for the experimental demonstration. The nanowire network was formed on a GaAs/AlGaAs heterostructure wafer by electron beam lithography and wet chemical etching. A thin SiN_x ($x = 1.2$) charge trapping layer was inserted between the WPG and the nanowire. The programming is carried out by charging and discharging the traps. This structure is simple and no special process is necessary to form. The SiN_x layer was deposited by electron cyclotron resonance chemical vapor deposition (ECR-CVD). The SiN_x thickness was 20 nm. Figure 5.13b shows an example of the hysteresis characteristic. Charging or discharging the electron traps by applying positive or negative gate voltage, respectively, switches the nanowire conductance on and off. When the gate voltage is back

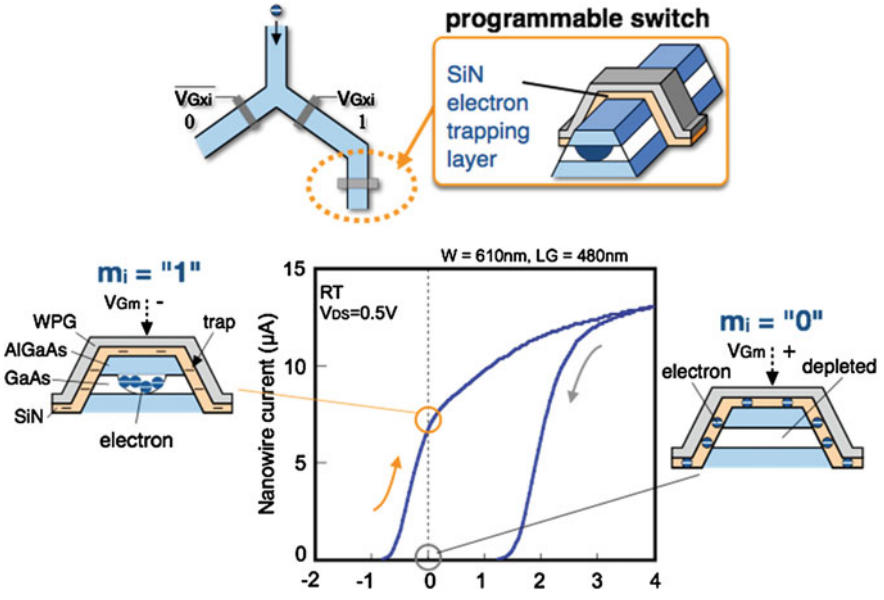


Fig. 5.13 Schematic illustration of the programmable switch using metal/SiN/GaAs nanowire structure and example of the hysteresis characteristic

to 0 V, it holds the previous state. Figure 5.14 shows the dynamic programming and switching characteristics of a fabricated switch. The initial threshold voltage before programming was −0.9 V. The on and off states were programmed by applying −2 and 2 V to the gate, respectively, for 1 ms. The switch held the conductance states for 200 ms, which was shorter than that of the conventional Si-based charge trap memory [55]. However, it was enough to demonstrate the feasibility of the circuit if the clock speed is sufficiently higher than the inverse of the retention time. From the measured hysteresis curve, the threshold voltage shift was 1.4 V and the trap density was estimated to be 2.5×10^{12} cm^{-2}. The retention time was prolonged by increasing either programming voltage V_{Gm} or programming time. Inserting a barrier layer between the SiN_x and the nanowire will improve the retention time characteristics.

A SEM image of a fabricated two-input reconfigurable BDD circuit is shown in Fig. 5.15 with its schematic. The width and height of the rectangle bounding a hexagon is 4.5 and 6.7 μm, respectively, and the total circuit area is 13×21 μm^2. The fabrication process used for this study achieves the node density of 4 M nodes/cm^2. The nanowire width, W, is 610 nm. The gate lengths, L_G, of the node devices and programmable switches are 380 and 480 nm, respectively. In this circuit, three node devices and four programmable switches are integrated.

Measured input-output waveforms of the fabricated circuit are shown in Fig. 5.16. An ideal output waveform is also shown in this diagram. The circuit was successively programmed to operate AND, NAND, OR, NOR, XOR, and XNOR functions. Supply voltage, V_{DD}, was −0.5 V. The input voltage swing was 0.6 V, and the offset voltage was 0 V for all inputs. The input signal frequency was 100 Hz. The current in the common terminal was measured as output. In accordance with the table in Fig. 5.12a, m_j was programmed by applying −2 V for "1" and 2 V for "0" to the programmable switches' gate. Each programming time was 1 ms. The program was changed sequentially at intervals of 200 ms. Correct output for each programmed function was successfully obtained, and dynamic reconfiguration was achieved in the

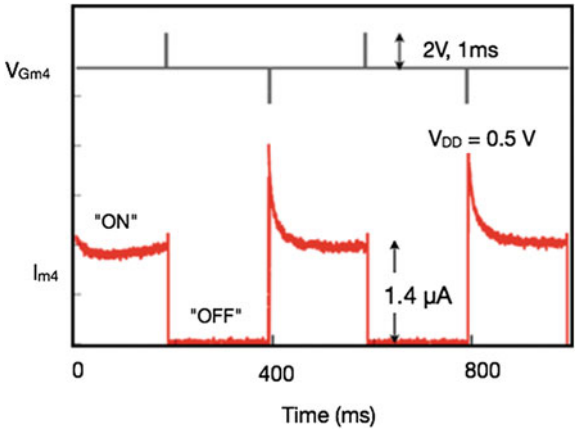


Fig. 5.14 Dynamic programming and switching characteristics of a fabricated programmable switch

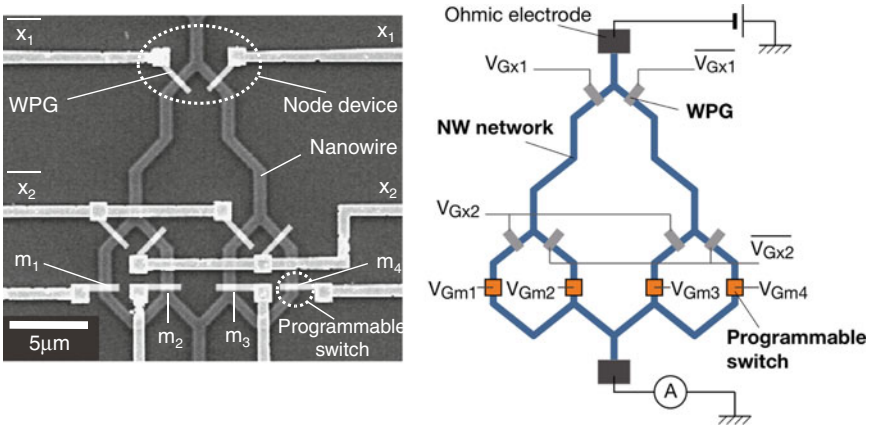


Fig. 5.15 Fabricated 2-input reconfigurable BDD circuit with its schematics

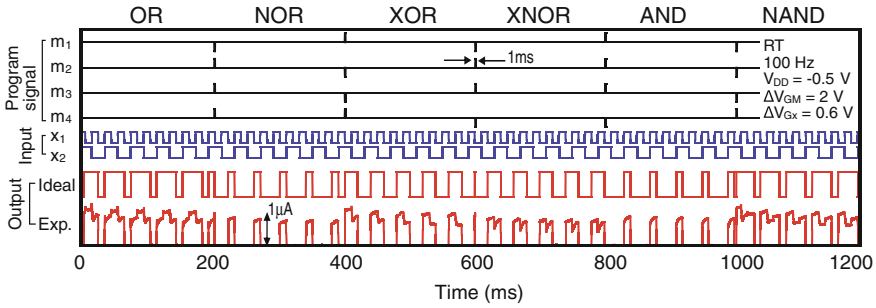


Fig. 5.16 Measured input-output waveforms with dynamic programming

fabricated circuit. These results demonstrate the feasibility of the proposed reconfigurable BDD circuit.

There was fluctuation in the measured output current. Owing to the graphical architecture, we could identify the origin of the fluctuation easily. This is a clear advantage of the circuit. For example, step-like fluctuation such as that in OR, XOR, and NAND functions was caused by the coupling with interconnects and nanowires for x_2 and $\bar{x}_2$. The delay was caused by the unintentionally slower switching characteristic of the WPG on the right branch of the node device labeled x_1. Such delay has been sometimes observed in GaAs-based FETs due to surface states in the gate periphery [56, 57].

5.3.3 Implemented Functions and Circuit Aarea

One of the advantages of the proposed circuit is compactness: a small number of devices and a small area. This is resulted from a compact data structure of the BDD [4]. For example, the proposed circuit integrates a total of 7 devices (10 WPGs) for two-input functions, whereas 42 transistors are required in a two-input look-up table for a standard Si CMOS design [58]. The area of the proposed circuit depends on the density of the nanowire network. A high-density fabrication process for 45 M nodes/cm^2 is available [41], which reduces the area to 9 % of the fabricated circuit in Fig. 5.15. The area also depends on the number of implemented logic functions. For example, limiting the number to half by choosing logics in which $m_2 = m_3$ would reduce the area of this circuit by 50 % by using a reduced order BDD technique [4]. Since the useful logic functions are symmetric, where the output is the same when the values of x_1 and x_2 are exchanged, the half of the functions in the full tree is not used so frequently. The number of programmable switches is also reduced from four to three. This technique will be pronounced in more large-scale circuits.

The topology of the network affects on not only the circuit area but also the wiring for gate inputs. Figure 5.17 shows fabricated 4-inputs reconfigurable BDD circuits with hexagonal and hyper-cube topology. Each circuit integrates 15 node devices and 16 programmable switches. $2^{16} = 65{,}536$ functions can be computed in one circuit. The estimated circuit areas for the two topologies are plotted in Fig. 5.18. Owing to higher asymmetry, the hypercube gives the smaller area in a smaller bit sizes. However the hexagonal layout provides smaller area when the bit size is large, because interconnects turn to occupy the area and to be dominant.

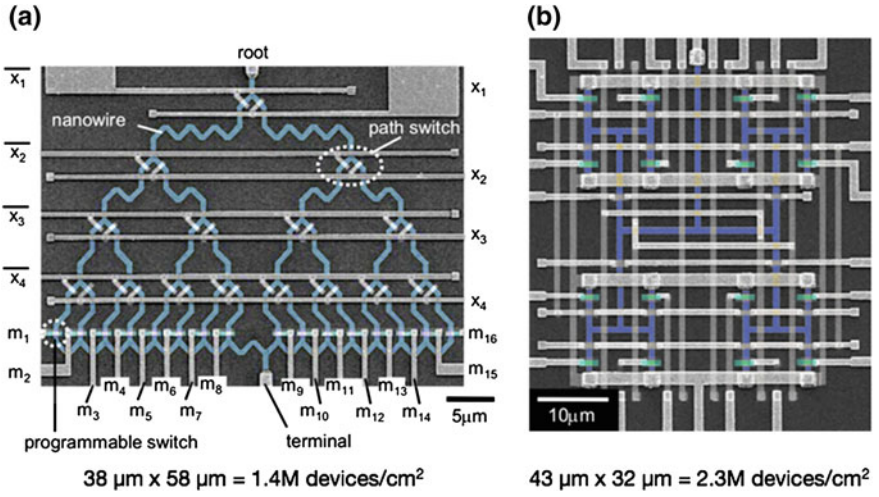


Fig. 5.17 Fabricated 4-input reconfigurable circuits with **a** hexagonal and **b** hypercube topologies

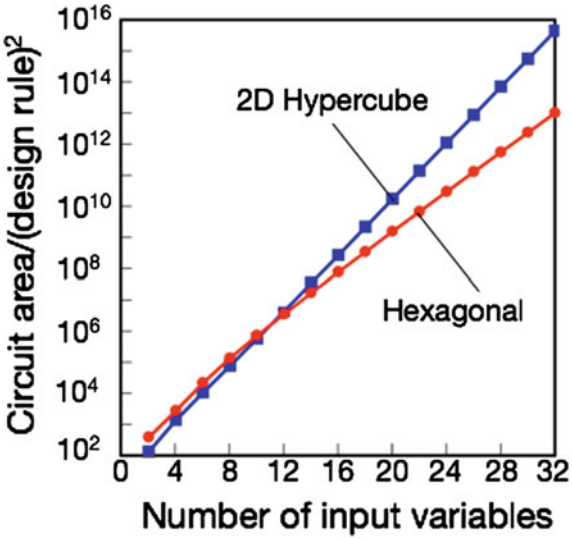


Fig. 5.18 Estimated circuit area versus number of input variables

5.4 Error Correcting BDD Circuits

5.4.1 Background

In this section, the error correcting BDD circuit is introduced. When the device size is reduced, it is going to suffer from more errors, because the signal is small and also it becomes sensitive to the environmental. Low voltage and low current in which a few electrons carry information are signals that are more susceptible to noise and thermal variations. Therefore, noise tolerance becomes an important issue. Another source for error is incorrect switching at the nodes or missing wiring due to defects [59]. Techniques to overcome the incorrect operation of circuits have been studied for a long time [60, 61]. Stochastic properties of circuits have been considered by many researchers by modeling the circuits as a communication system with a noisy channel [62]. We apply error-correction techniques from communication theory to design logic circuits. The goal is not to restore the values of the variables; the binary message. The goal is to keep the output of the binary function correct. Since there are several combinations of values that cause the binary function to output the same result, changing the input values to any of them will not result in an error. This is different from the usual context of error correction encoding and decoding, because now the properties (truth table) of the function itself are important. Applying the theory of block codes in this context is straightforward. Hamming codes are a subset of block codes and they have the desired property of correcting a single error using the minimum number of extra parity bits. They are suitable for small binary messages with minimum coding overhead. The suitability of cyclic codes needs to be

further studied. Cyclic codes in general have the advantage over block codes of easier encoding and decoding which is not our target. Convolutional codes are suitable for continuous data streams and they target retrieving the message. This is against the requirement that the groups of binary messages representing the function inputs are applied one at a time. There is also no direct relation between the BDDs and trellis decoders. However, this approach has not been implemented at circuit level yet. In this section, the error correction capability is demonstrated by a simulation of the circuit using a conventional circuit simulator. We analyze their behavior under the effect of both noise and random signal propagation errors. The structure of such circuits corresponds to the nanowire-network BDD circuits [12]. Therefore, we consider such modeling as a prerequisite to prototyping and manufacturing the nanostructure-based circuits with error-correction capability.

5.4.2 Error Correction Scheme

One of the widely used ways to detect/correct error in the communication systems is parity error check/correction. Redundant bits are added to the original bit signal. By the help of the redundant bit (block code), error in the output is detected or corrected. Astola et al. suggests combining the block error correcting codes and decision diagrams [63]. A block code, denoted as (n, k), consists of n-digit codewords with $n > k$, which map onto a smaller set of k digit words. A block code is linear, if the modulo-2, sum of any two codewords is also a codeword. The Hamming distance between two codewords is the number of digits, in which they are different from each other. This can be used error detection or both error detection and correction. The latter can be performed when the Hamming distance gives enough separation to determine which codeword is the most likely one. For example, if the Hamming distance is 3, it is possible to correct one error and detect two errors. The codewords are generated using a generator matrix. The original codeword is restored using a parity check matrix [64].

For a logic function f of k variables, its original BDD includes k levels. If a switching error happens at a BDD node, it cannot be corrected. The error-correcting BDD is derived by extending the BDD: a binary code n, k is constructed, and the function f is mapped into a function that is implemented using an n-level BDD [63]. The error correction functionality is embedded in the BDD circuitry and the circuit itself corrects or detects the error. Note that the increase in BDD size is of linear complexity and a small price is paid for error correction. A typical system is shown in Fig. 5.19. The system consists a parity bit generating BDD and an error-correcting BDD. In the beginning, a codeword including parity bits is generated by operating the input on the parity generating BDD that represents a parity bit generator matrix. Then the codeword is given to the error correcting BDD that evaluating the logic function. This BDD gives the output of the original logic function.

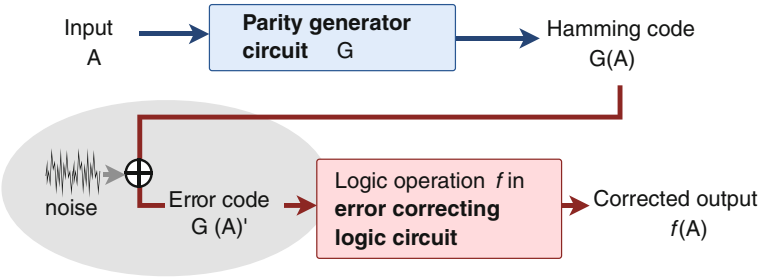


Fig. 5.19 Basic concept of the error correcting BDD

5.4.3 Implementation and Characterization

One Variable Buffer

Consider a buffer circuit as an example, whose output signal is the same to the input. It is a one-level BDD ($k = 1$) as shown in Fig. 5.20a. The function of one variable corresponds to the (3,1) error-correcting code. The parity generating matrix is $\boldsymbol{G} = [111]$. Operating one variable to this matrix, "0" is encoded by 000 and "1" is encoded by 111 [65]. A decoder of such a code represents a majority-vote function: it decodes the received codewords 000, 001, 010 and 100 as "0", and 111, 110, 101, and 011 as "1". Therefore, it can correct one bit in the corresponding codewords. The error-correcting BDD of the buffer has three levels as shown in Fig. 5.20c. This BDD is derived from a complete three-level binary tree, using the reduction rules. Even if a switching error happens at one level of the BDD, the output value of the function can be still evaluated correctly. A NOT gate can be represented by exchanging 0 and 1 terminals of the Buffer BDD as shown in Fig. 5.20b. The error-correcting code and the generating matrix are the same. Figure 5.21 shows the results of a circuit simulation. The response of the circuit suffered from large superimposed noise at one of the switching levels as shown in Fig. 5.21a. The noise component is

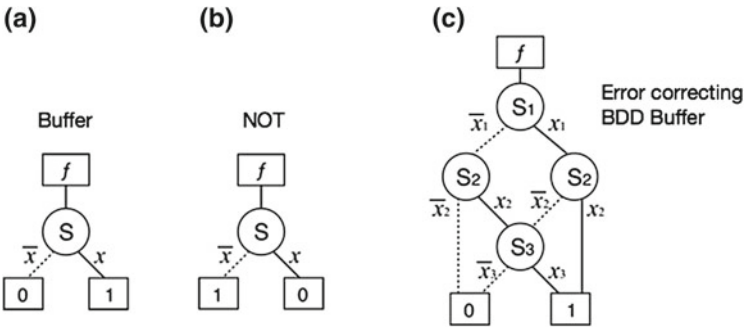


Fig. 5.20 **a** Normal BDD buffer, **b** BDD NOT gate, and **c** error correcting BDD buffer

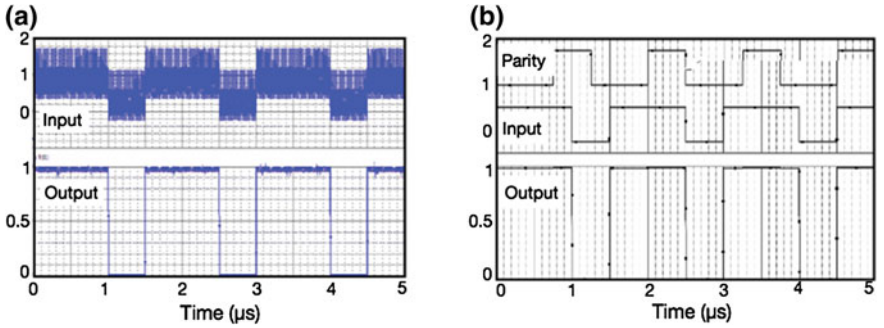


Fig. 5.21 **a** Response of buffer to noisy input and **b** response of the buffer to unintentionally phase shifted signal

greatly suppressed in the output and the correct response to the original input signal is realized. Figure 5.21b demonstrates that an incorrect signal propagation delay between two nodes does not affect the operation and the correct output is obtained.

Fundamental Two-Input Logic Gates

Any 2-input fundamental logic gate, such as AND, NAND, OR, NOR, EXOR or XNOR, can be implemented by a two-level binary decision tree. For example, a binary tree and BDD for the NAND gate is shown in Fig. 5.22a. The Hamming codes, corresponding to the 2-input logic functions, can be a perfect (7, 4) code, or a non-perfect (2, 5) code. The generating matrix for the non-perfect (2, 5) code is given by **G** in Fig. 5.22b. The parity generator BDD is also shown in the same figure. The code generated by this matrix is non-perfect, because two errors yield an undefined output value (which can be treated as 0, 1 or a different value). The corresponding error-correcting BDD has five levels. Figure 5.22c shows the design of the error-correcting BDD for NAND derived from a five-level binary tree.

The result of simulation of the NAND gate is shown in Fig. 5.23a, when the switching nodes of two inputs are affected by noise with large amplitude comparable to the control signal amplitude. This demonstrates the strong error correction and better tolerance against the dynamic random signal of the present circuit. Note that the output is almost perfectly recovered even with two uncorrelated noises are given to the two input. Considering that the circuit cannot correct two bit errors, the result suggests that the circuit almost suffers from one bit error in each time. Thus one bit correction with the non-perfect code is enough for the two-bit logic gates for certain noises, whereas this property depends on correlation between noises in the set of inputs. Figure 5.23b plots the measured bit error rate (BER) as a function on input signal-to-noise ration (SNR). BDD-based error correcting circuits always shows lower BER than the conventional BDD.

Subsystem

We examine the present approach in the subsystem level. Figure 5.24 shows the simulated input-output waveforms of the 2-bit adders designed by the conventional

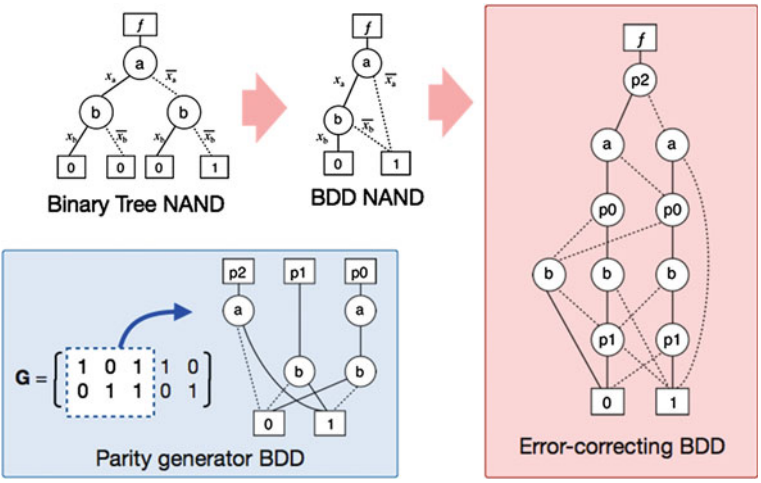


Fig. 5.22 **a** Binary tree and BDD of 2-input NAND, **b** parity generating matrix and generator BDD, and **c** error-correcting BDD of the NAND

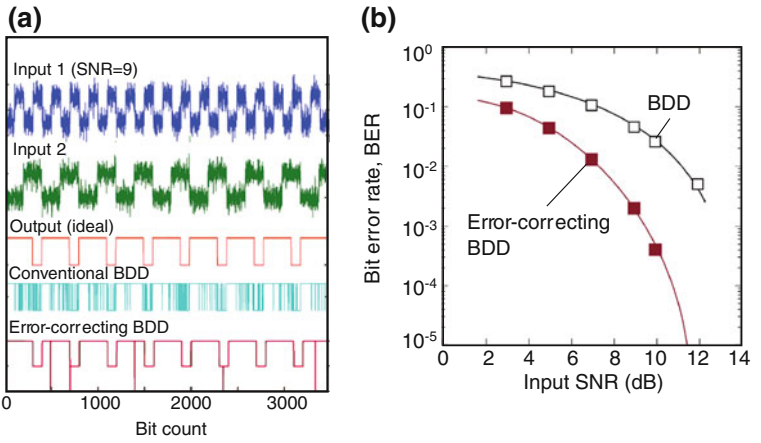


Fig. 5.23 **a** Simulated input-output waveforms of the BDD circuits without and with error correction and **b** measured BER versus. input SNR

BDD and the error-correcting BDD. Noisy inputs are given to four inputs, a_1, a_0, b_1, and b_0. The error correcting BDD uses (7, 4) error-correcting code, that is 4 data bits and 3 parity bits. The parity bit generator with 4-levels consists of 9 node devices. The adder circuit consists of 68 devices, whereas that of the normal BDD consists of 11 devices. We have already developed tools for automatically generating codes and BDD. Without correction, all the outputs are almost random as shown in Fig. 5.24a. However, the error correcting BDD circuits almost correct outputs although a small number of spike errors appear. Measured BER in the error correcting BDD again

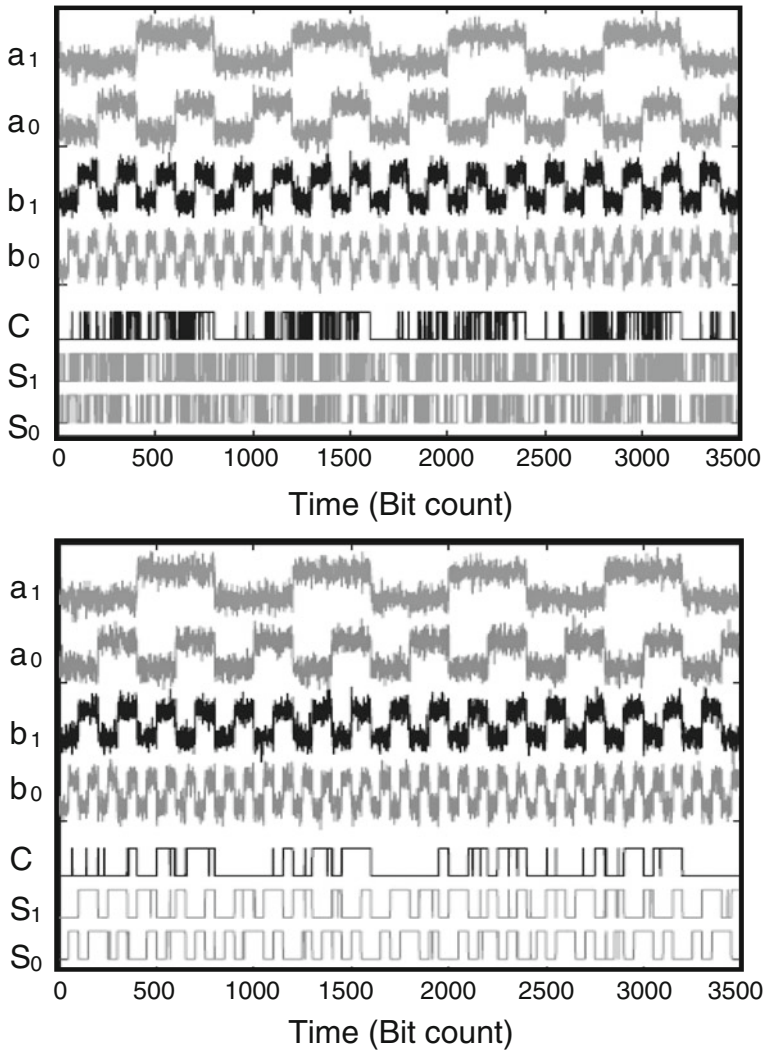


Fig. 5.24 Simulated input-output waveforms of 2-bit adders by **a** conventional BDD and **b** error-correcting BDD with noisy inputs (SNR = 9 dB)

gives better error rate than that of the conventional BDD in all input SNR range. The obtained results confirm the good feasibility of the error correcting BDD. In order to realize a circuit using a regular hexagonal BDD array, it is mandatory to modify the circuit so that it becomes planar [66]. This can be achieved by the insertion of dummy nodes, node swapping or duplication, and control signal duplication.

5.5 Summary

Boolean logic circuits utilizing nanowire networks and related technologies based on graphical representation of logic functions using a binary decision diagram (BDD) were presented. The physical network configuration itself represents a logic function as a data structure, unlike the conventional logic circuit in which each logic gate has a specific function as an operator.

Feasibility of the nanowire-network-based BDD circuit was demonstrated through fabrication and characterization of a two-bit arithmetic logic unit (ALU) on a GaAs-based regular nanowire network with hexagonal topology. A four-instruction two-bit ALU was designed by integrating subgraphs representing each instruction. The fabricated circuit integrating 32 node devices operated correctly at room temperature, allowing for threshold voltage variation. The obtained results indicated that the present circuit approach can possibly utilize various nanodevices that are difficult to use in current-driven integrated circuits having logic gate architecture.

A reconfigurable BDD logic circuit based on Shannon's expansion and its graphical representation on a nanowire network is presented. This circuit architecture has a simpler structure and is more compact than the previous reconfigurable BDD and conventional CMOS look up table. The two-input reconfigurable BDD circuit was fabricated on an etched GaAs hexagonal nanowire network with Schottky wrap gates and SiN_x-based programmable switches. Dynamic reconfiguration and correct logic operation were obtained in the fabricated circuit by suitable programming.

Fault-tolerant design concept of such circuits, using the binary decision diagram (BDD) and error-correcting techniques was described. The preliminary simulation results for incorporating error correction in switching nanowire circuits using linear block codes showed that it was feasible to design such circuits on nanowire network-based BDD, via introduction of linear-complexity redundancy to the circuit. The simulation confirmed that such structures combat errors, caused by noise in a signal, or by the switching errors and wire defects.

Acknowledgments This work is partly supported by Grants-in-Aid for Scientific Research from Japan Society for Promotion of Science (JSPS), Japan and by Natural Sciences and Engineering Research Council (NSERC) of Canada. Dr. Yanushkevich and Dr. Kasai acknowledge JSPS Invited Fellowship program. Mr. T. Mohamed acknowledges the iCore program (Alberta, Canada).

References

1. International Technology Roadmap for Semiconductors 2011 edition, Emerging Research Materials (http://www.itrs.net/)
2. M. Bohr, presented at Microprocessor. Forum **2007** (May), 21–23 (2007)
3. S.B. Akers, IEEE Trans. Compt. **C-27**, 509 (1978)
4. S.N. Yanushkevich, D.M. Miller, V.P. Shmerko, R. S. Stankovic, *Decision Diagram Techniques for Micro-and Nanoelectronic Design Handbook*. (CRC Press, Florida, 2006)
5. Y. Chen, G.-Y. Jung, D.A.A. Ohlberg, X. Li, D.R. Stewart, J.O. Jeppesen, K.A. Nielsen, J.F. Stoddart, R.S. Williams, Nanotechnology **14**, 462 (2003)

6. K. Terabe, T. Hasegawa, T. Nakayama, M. Aono, Nature **433**, 47 (2005)
7. Y. Naitoh et al., Nanotechnology **17**, 2406 (2006)
8. G. Snider, P. Kuekes, R.S. Williams, Nanotechnology **15**, 881 (2004)
9. H. Hasegawa, S. Kasai, Physica E **11**, 149 (2001)
10. S. Kasai, H. Hasegawa, IEEE Electron. Device Lett. **23**, 446 (2002)
11. S. Kasai, T. Nakamura, Y. Shiratori, T. Tamura, J. Comp. Theoretical Nanosci. **4**, 1120 (2007)
12. H.-Q. Zhao, S. Kasai, Y. Shiratori, T. Hashizume, Nanotechnology **20**, 245203 (2009)
13. S. Kasai, M. Yumoto, H. Hasegawa, Solid State Electron. **47**, 199 (2003)
14. S. Kasai, M. Yumoto, T. Sato, H. Hasegawa, ECS Proc. **2004–13**, 125 (2004)
15. N. Yoshikawa, H. Tago, K. Yoneyama, IEEE Trans. Appl. Supercond. **9**, 3161 (1999)
16. N. Yoshikawa, F. Matsuzaki, N. Nakajima, K. Fujiwara, K. Yoda, K. Kawasaki, IEEE Trans. Appl. Supercond. **13**, 441 (2003)
17. M. Yumoto, S. Kasai, H. Hasegawa, Microelectron. Eng. **63**, 287 (2002)
18. N. Asahi, M. Akazawa, Y. Amemiya, IEEE Trans. Electron. Devices **42**, 1999 (1995)
19. Y. Shiratori, K. Miura, S. Kasai, R. Jia, N.-J. Wu, S. Kasai, Appl. Phys. Express **3**, 025002 (2010)
20. V.P. Shmerko, S.N. Yanushkevich, S.E. Lyshevski, *Computer Arithmetics for Nanoelectronics* (Taylor & Francis/CRC Press, Boca Raton, 2009)
21. K. Yano, T. Yamanaka, T. Nishida, M. Saito, K. Shimohigashi, A. Shimizu, IEEE J. Solid State Circ. **25**, 388 (1990)
22. Q. Lu, F. Gao, S. Komarneni, J. Am. Chem. Soc. **126**, 54 (2004)
23. H. Hasegawa, T. Sato, S. Kasai, in *Lateral Alighment of Epitaxial Quantum Dots*, ed. by O.G. Schmidt (Springer, New York, 2007), pp. 637–664
24. H. Yan, R. He, J. Johnson, M. Law, R.J. Saykally, P. Yang, J. Am. Chem. Soc. **125**, 4728 (2003)
25. K. Kumakura, J. Motohisa, T. Fukui, J. Cryst. Growth **170**, 700 (1997)
26. S. Nakanishi, K. Fukami, T. Tada, Y. Nakato, J. Am. Chem. Soc. **126**, 9556 (2004)
27. J. Zhou, Y. Ding, S.Z. Deng, L. Gong, N.S. Xu, Z.K. Wang, Adv. Mater. **17**, 2107 (2005)
28. T. Nakamura, Y. Abe, S. Kasai, H. Hasegawa, T. Hashizume, J. Physics, Conf. Ser. **38**, 104 (2006)
29. W. Hioe, M. Hosoya, S. Kominami, H. Yamada, R. Mita, K. Takagi, IEEE Trans. Appl. Supercond. **5**, 2992 (1995)
30. S. Minato, N. Ishiura, S. Yajima, in *Proceedings 27th ACM/IEEE Design Automation Conference*. (Orlando, 1990), p. 52. 24–28, June 1990
31. T. Kuroda, T. Sakurai, IEICE Trans. Electron. **E78-C**, 334 (1995).
32. A. Naeemi, Y. Joshi, A. Fedorov, P. Kohl, J.D. Meindl, in *2005 International Conference on Integrated Circuit Design and Technology*. (Padova, 2005), p. 171. 9–11 May 2005
33. M. Yumoto, S. Kasai, H. Hasegawa, IOP Conf. Ser. **184**, 213 (2005)
34. H.-Q. Zhao, S. Kasai, T. Hashizume, N.-J. Wu, IEICE Trans. Electron. **E91-C**, 1063 (2008)
35. H. Shibata, Y. Shiratori, S. Kasai, Jpn. J. Appl. Phys. **50**, 06GF03 (2011)
36. Y. Nakano, T. Tanaka, S. Kasai, in *Presented at International Workshop on Quantum Nanostructures and Nanoelectronics (QNN2011)*. (Tokyo, 2011). 3–4 October 2011
37. M. Yumoto, S. Kasai, H. Hasegawa, Jpn. J. Appl. Phys. **41**, 2671 (2002)
38. M. Yumoto, S. Kasai, H. Hasegawa, Appl. Surf. Sci. **190**, 242 (2002)
39. H. Kawaura, T. Sakamoto, T. Baba, *1999 Si Nanoelectronics Workshop Abstracts*. (Kyoto, 1999), p. 26. 10–11 June 1999
40. Y. Shiratori, S. Kasai, Jpn. J. Appl. Phys. **48**, 3086 (2008)
41. S. Kasai, M. Yumoto, T. Fukushi, T. Muranaka H. Hasegawa, in *Digest of 60th Annual Device Research Conference*. (Santa Barbara, 2002), p. 103. 24–26 June 2001
42. O. Karlström, A. Wacker, K. Nilsson, G. Astromskas, S. Roddaro, L. Samuelson, L.-E. Wernersson, Nanotechnology **19**, 435201 (2008)
43. K. Tomioka, P. Mohan, J. Noborisaka, S. Hara, J. Motohisa, T. Fukui, J. Cryst. Growth **298**, 644 (2007)
44. K.A. Dick, K. Deppert, L.S. Karlsson, W. Seifert, L.R. Wallenberg, L. Samuelson, Nano Lett. **6**, 2842 (2006)

45. H. Soeleman, K. Roy, in *Proceedings of 1999 International Symposium on Low Power Electronics and Design*. (San Diego, 1999), p. 94. 10–17 August 1999
46. T. Nakamura, Y. Shiratori, S. Kasai, T. Hashizume, Appl. Phys. Lett. **90**, 102104 (2007)
47. S. Kasai, T. Nakamura, S.F. Bin Abd Rahman, Y. Shiratori, Jpn. J. Appl. Phys. **47**, 4958 (2008)
48. A.R. Shaharin Fadzli, D. Nakata, Y. Shiratori, S. Kasai, Jpn. J. Appl. Phys. **48**, 06FD01 (2009)
49. D. Nakata, H. Shibata, Y. Shiratori, S. Kasai, Jpn. J. Appl. Phys. **49**, 06GG03 (2010)
50. H.Q. Xu, Appl. Phys. Lett. **78**, 2064 (2001)
51. S. Kasai, T. Asai, Appl. Phys. Express **1**, 083001 (2008)
52. S. Kasai, K. Miura, Y. Shiratori, Appl. Phys. Lett. **96**, 194102 (2010)
53. J.R. Heath, P.J. Kuekes, G.S. Snider, R.S. Williams, Science **280**, 1716 (1998)
54. S. Eachempati, V. Saripalli, N. Vijaykrishnan, S. Datta, in *Proceedings IEEE/ACM International Symposium On Nanoscale Architectures*. (Amsterdam, 208), p. 61. 12–13 June 2008
55. S. Aritome, R. Shirota, K. Sakui, F. Masuoka, IEICE Trans. Electron. **E-77-C**, 1287 (1994)
56. H. Sasaki, H. Matsubayashi, O. Ishihara, R. Konishi, K. Ando, Jpn. J. Appl. Phys. **34**, 6346 (1995)
57. Y. Ohno, P. Francis, M. Nogome, Y. Takahashi, IEEE Trans. Electron. Devices **46**, 214 (1999)
58. T. Sueyoshi, H. Amano, *Rikonfigyaraburu shisutemu* (Reconfigurable System). (Ohmsha, 2005), p. 8 [in Japanese]
59. P.J. Kuekes, W. Robinett, R.S. Williams, Nanotechnology **17**, 2466 (2006)
60. A.H. Tran, S.N. Yanushkevich, S. Lyshevski, V. Shmerko, in *Proceedings 11th IEEE International Conference Nanotechnology*. (Portland, 2011), p. 457. 15–18 Aug 2011
61. S. Winograd, J.D. Cowan, *Reliable Computation in the Presence of Noise* (MIT Press, Cambridge, 1963)
62. G.V. Varatkar, S. Narayanan, N.R. Shanbhag, D.L. Jones, IEEE Trans. VLSI Syst. **18**, 1421 (2010)
63. H. Astola, S. Stankovic, J.T. Astola, in *Proceedings of 3rd Workshop on Information Theoretic Methods in Science and Engineering*. (Tampere, 2010), 16–18 August
64. J.G. Proakis, M. Salehi, *Fundamentals of Communication Systems* (Pearson Prentice Hall, New Jersey, 2005)
65. F.J. MacWilliams, N.J.A. Sloane, *The Theory of Error-Correcting Codes* (North-Holland, Amsterdam, 1997)
66. A. Cao, C.K. Koh, in *Proceedings 21st International Conference Computer Design (ICCD)*. (San Jose, 2003), p. 338. 13–15 October 2003

Chapter 6
Single Photoelectron Manipulation and Detection with Sub-Nanosecond Resolution in CMOS Imagers

Shoji Kawahito, Keita Yasutomi and Keiichiro Kagawa

Abstract This chapter describes single photoelectron manipulation and detection with sub-nanosecond time resolution in CMOS imagers. Base on an lateral electric field control in a perfectly depleted photodiode without any potential barrier, a single photoelectron transfer in less than one nano second is possible. This property is particularly useful for time-resolved very low light level biological imaging such as in fluorescent lifetime imaging.

6.1 Introduction

Complementary metal-oxide-semiconductor (CMOS) image sensors (CISs) have replaced charge-coupled device (CCD) image sensors as a major imaging device because of their excellent performance, including high image quality, high resolution, and high frame rate. Before 2010, high image quality and high photosensitivity were CCD advantages. However, due to the development of the four-transistor (4T) pixel structure based on a pinned photodiode (PPD) and low-noise column parallel amplifiers or column-parallel analog-to-digital converters (ADCs), more CMOS image sensors than CCD image sensors are being used. Furthermore, introduction of back-side illumination to CMOS image sensors has enabled more even high quantum efficiency. Low cost is another advantage of CMOS image sensors in mass production. In addition, standard CMOS image sensor technologies have become available to image-sensing researchers and analog or mixed-signal engineers, who thus

S. Kawahito (✉) · K. Yasutomi · K. Kagawa
Research Institute of Electronics, Shizuoka University, Hamamatsu, Japan
e-mail: kawahito@idl.rie.shizuoka.ac.jp

K. Yasutomi
e-mail: kyasu@idl.rie.shizuoka.ac.jp

K. Kagawa
e-mail: kagawa@idl.rie.shizuoka.ac.jp

M. Naruse (ed.), *Nanophotonic Information Physics*,
Nano-Optics and Nanophotonics, DOI: 10.1007/978-3-642-40224-1_6,

gain the capability to design a dedicated pixel and a functional or high-performance analog front-end or mixed-signal circuit. Because CCD technology is very different from standard CMOS technology, such capability is not available to everyone. Nevertheless, many researchers and developers have gained a strong tool to realize high-performance and high-quality dedicated image sensors. It can be said that the era of tailor-made image sensors has begun.

Biological imaging is a very important application field for dedicated CMOS image sensors. The requirements for biological imaging are often very different from those for ordinary consumer digital cameras. For example, biological imaging requires extremely low light (a few photons per frame), ultra-high speed (shorter than several nanoseconds), and extremely low contrast. In fluorescence lifetime imaging microscopy (FLIM) [1], tens of picoseconds to sub-nanoseconds of temporal resolution is required, and in stimulated Raman spectroscopy (SRS) [2], contrast of the optical signal is typically less than 1:10000. The fundamental technology for CMOS image sensors to capture these extreme images is single photoelectron manipulation at ultra-high speed without any loss and delay. In fact, the time of electron transfer in one micrometer under the electric field of 1 V/μm in depleted silicon is a few tens of picoseconds. If the potential distribution of the channel region is elaborately designed by metal-oxide-semiconductor (MOS) gates without any potential barriers in the channel, the two-dimensional electric field can be controlled by the gates. Hence, the behavior of the single photoelectron in the time scale of nanoseconds can be manipulated. This property is particularly useful for time-resolved biological imaging at a very low light level. This chapter describes single photoelectron manipulation and detection with sub-nanosecond time resolution in a pinned diode of CMOS imagers and a practical application.

6.2 Single Photoelectron Manipulation by CMOS Active Pixel

Two approaches are possible for single photoelectron manipulation in a pixel of CMOS imagers: avalanche photoelectron multiplication [3] and photoelectron flow control based on two-dimensional electric field modulation [4]. In either approach, a pixel containing active circuits for amplification or signal processing is called an active pixel. First, these two approaches are compared. Then, we focus on photoelectron flow control.

6.2.1 Single Photon Avalanche Diode

A promising event-driven photon detection scheme is based on a single photon avalanche diode (SPAD) that operates in Geiger mode. Figure 6.1 depicts a photodiode structure and a schematic drawing of electron multiplication. The p-n junction is reversely biased at a high voltage that is slightly lower than the avalanche breakdown

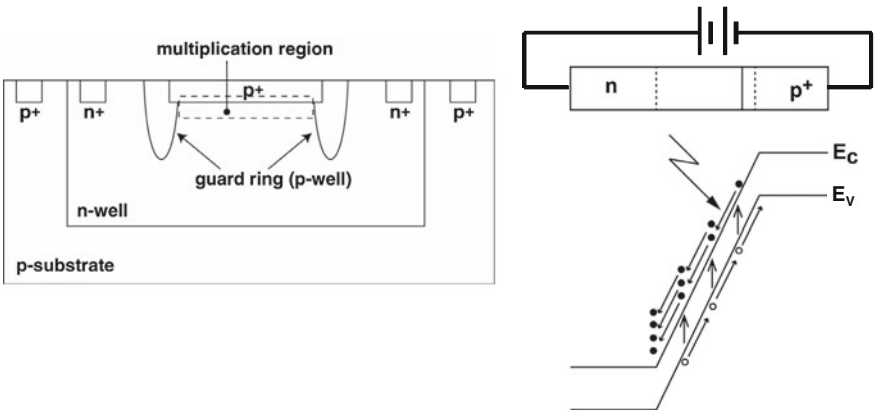


Fig. 6.1 A structure of SPAD and photoelectron multiplication

voltage. The cathode voltage of the photodiode is V_{op} at this time. A photon absorbed in the depleted region triggers an avalanche multiplication, so that many multiplied electrons are generated from a single photon. Figure 6.2 shows a basic photodetection circuit of a SPAD. Electrons multiplied by the avalanche breakdown readily decrease the p-n junction voltage to zero, and the cathode voltage of the SPAD almost reaches V_{bd}. Then, the SPAD is recharged to V_{op} by the PMOS transistor. This process is called quenching. The CMOS inverter reshapes the cathode voltage to a pulse. Thus, an event of photon arrival is represented by a digital pulse.

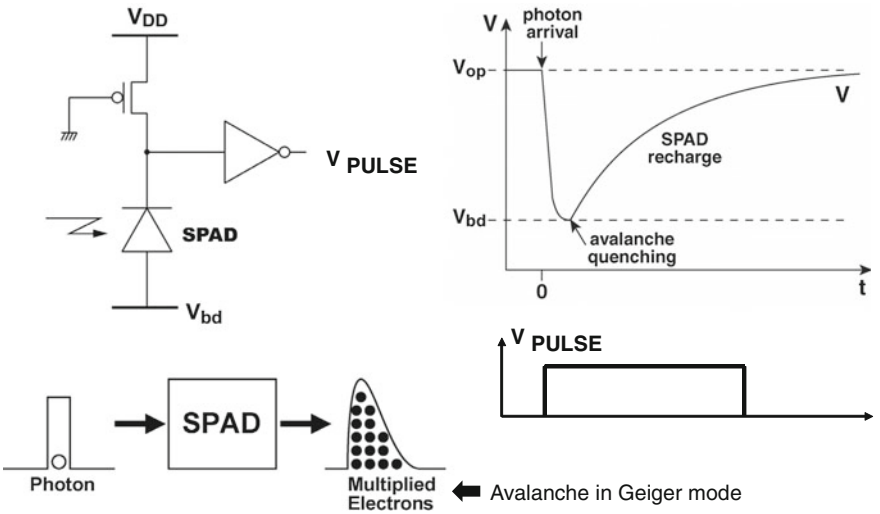


Fig. 6.2 Basic detection circuit of SPAD and its operation

In the SPAD, every photon is identified as an independent pulse and can be time-stamped. Typically, the photodetection circuit in Fig. 6.2 is followed by a time to digital converter (TDC), which enables measurement of the interval time between two adjacent events or measurement of the delay from the excitation short light pulse. A histogram analysis of the measured data shows the fast temporal behavior of the light signal with sub-nanosecond resolution.

The drawbacks of the SPAD are (1) low quantum efficiency, (2) large dark current (observed as dark count), (3) upper limit of the count rate, and (4) large area for detection circuits. Issues 1 and 2 are being addressed by process improvements. Issue 3 is caused by the quenching time, that is, the recovery time of the cathode voltage from V_{bd} to V_{op} after the avalanche multiplication. The quenching time cannot be shortened easily. Therefore, the peak light intensity is limited. As for Issue 4, although the area for the TDC and the signal processing circuits can be reduced by a finer process, it cannot be negligible. Hence, it is not easy to implement high-resolution image sensors based on the SPAD.

6.2.2 Achieving High Quality in CMOS Image Sensors

Before describing our single photoelectron manipulation approach, the technique of achieving high image quality in CMOS image sensors is worth noting because our pixel structure is dependent on this technique. A major contributor to image quality improvement in CMOS image sensors is the use of a PPD in a pixel, as shown in Fig. 6.3. Because of the buried photodiode structure, which does not have carrier

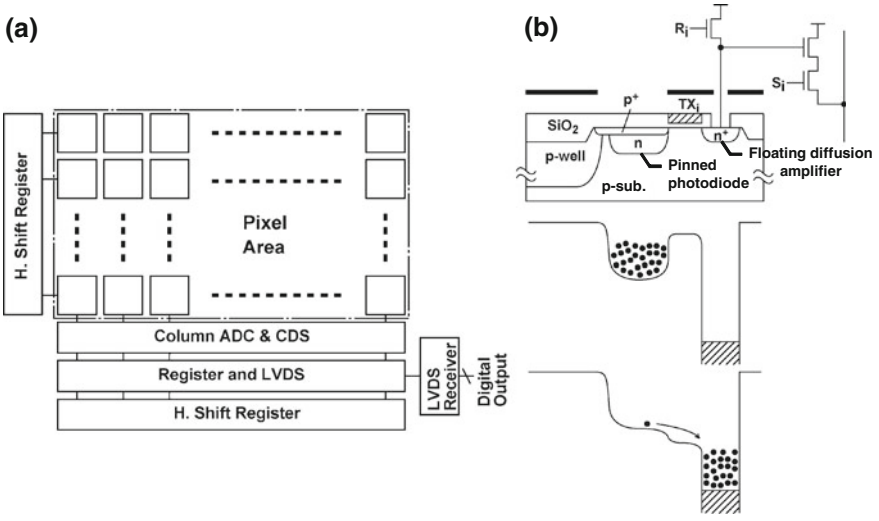


Fig. 6.3 **a** Block diagram of a typical CMOS image sensor and **b** 4T pixel structure with a pinned photodiode

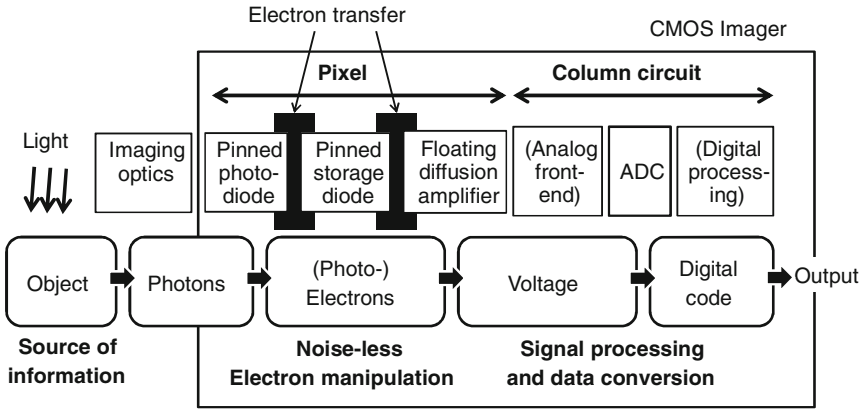


Fig. 6.4 Signal chain of our CMOS image sensors

trapping at the Si-SiO_2 interface and is fully depleted [5], high-speed photoelectron transfer in the photodiode is possible if the dimension, shape, and forming condition of the photodiode are well engineered. PPD also reduces the dark current significantly because traps at the Si-SiO_2 interface, which is a major dark current source, are filled with holes. The amount of photoelectrons is sensed at the floating diffusion (FD) amplifier after the photoelectrons in the PPD are completely transferred through the transfer gate (TX). Although reset noise or kTC noise is generated at the FD, the noise can be perfectly removed by a true correlated double sampling (CDS) operation. Figure 6.4 summarizes the signal chain in our CMOS image sensors. The pixel part is based on the 4T pixel. To realize charge modulation or time windowing, as well as CDS operation, a photodiode is followed by a storage diode. The diodes are implemented by a pinned diode to achieve high-speed electron transfer and low dark current. Photoelectron transfer between the PPD and the pinned storage diode (PSD) and between the PSD and the FD amplifier is controlled by MOS gates. Photoelectron transfer from the PPD to the PSD is the most crucial aspect in pixel design because it determines the electron transfer speed and efficiency. Hence, the potential shape from the PPD to the PSD should be engineered very carefully.

The analog front-end and the ADC technologies are as important as the pixel design. We have developed low-noise (approximately one electron rms) and high dynamic range ($>$80 dB) column ADCs. However, these are not the main points in this book. Refer to [6] for more information.

6.3 Our Method: Lateral Electric Field (LEF) Control

6.3.1 Concept and Implementation of LEF Control

Basic operations for photoelectron manipulation are sorting and draining. As discussed above, (1) noiseless, (2) high-speed (in less than one nanosecond), and (3) lossless photoelectron transfer are required for these operations. Requirement 1 can be satisfied by complete electron transfer with a pinned photodiode and a pinned storage diode. A fringing electric field can effectively satisfy Requirement 2, and any barrier should be avoided in the transfer path. To fulfill Requirement 3, any dip and $Si\text{-}SiO_2$ traps should be avoided because even a small dip traps a few photoelectrons so that no electrons are transferred in the low-light condition.

We have proposed lateral electric field (LEF) control, which is effective for noiseless, high-speed, lossless photoelectron transfer. The fundamental concept of LEF control is that the potential shape in the photoelectron transfer path is controlled by the potential on both sides of the path. There is a smooth fringing electric field in the transfer path, so neither a barrier nor a dip is generated in the path.

Figures 6.5 and 6.6 depict the basic structure of the LEF control pixel. The n_1 region is for the photodiode and the photoelectron transfer path, and the n_2 regions work as the storage diode. Both n_1 and n_2 regions are capped by a p+ layer. Control gates, G1 and G2, are located beside the n_1 region. Note that no n_1 region is underneath the control gates. The crucial design of this structure is that the p+ layer is softly pinned, although it is fixed to the p-substrate voltage (typically ground) in ordinary 4T pixels. Therefore, the voltage of the p+ layer can move slightly, which induces a potential change of the n_1 region adjacent to the control gate. As shown in Fig. 6.6, low-level voltage (−2 V) or high-level voltage (1 V) is applied to G1 and G2. When different voltages are applied to G1 and G2, a fringing electric field is generated in the n_1 region, as shown in Fig. 6.6a. Thus, the photoelectrons are readily

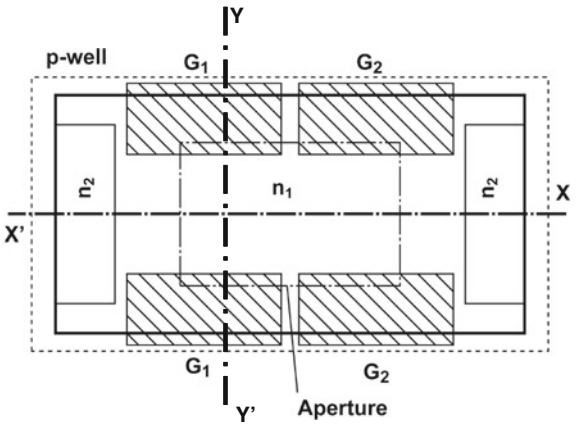


Fig. 6.5 Structure of LEF control pixel (*top view*)

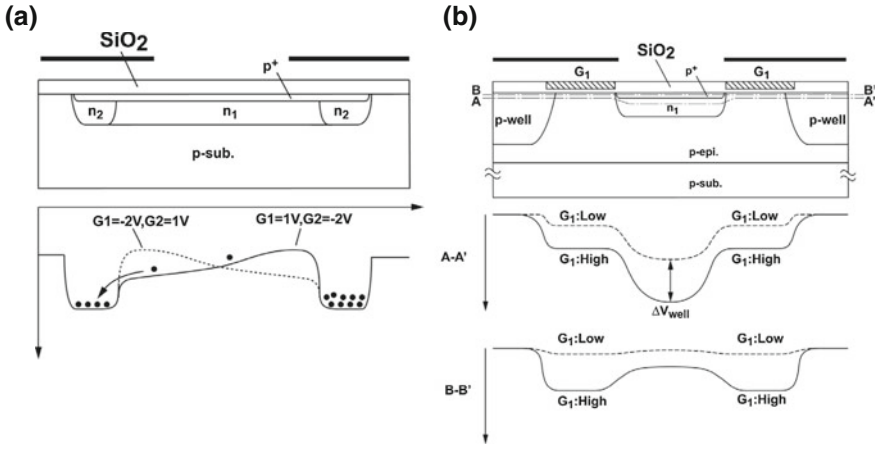


Fig. 6.6 Cross-sections of the LEF control pixel. **a** Cross-section and potential (X-X′), **b** Cross-section and potential (Y-Y′)

transferred to either location of the storage diodes. A concern is that dark current is generated under the control gates. However, the dark current can be alleviated by the accumulated holes beneath the control gates, which are generated while the control voltage is low.

For draining, a new gate, G3, and its diffusion region on the other side of the photodiode should be inserted between G1 and G2. The highest voltage (3.3 V) is applied to G3 while low-level voltage is applied to G1 and G2 to keep the photoelectrons in the storage diodes. Then, the photoelectrons at the photosensitive area are exhausted to the diffusion (drain).

Our detection approach is very different from that by a SPAD. In a SPAD, every photon is time-stamped. Therefore, pre-processing (e.g., time windowing) and post-processing (e.g., histogram analysis) are possible. On the other hand, our photoelectron modulation approach does not distinguish each photon. Pixels calculate the inner products of the temporal light intensity function and some base functions implemented by the control voltages applied to the gates in the photoelectron region, which is a noiseless operation. Typical base functions are repetitive rectangular functions with a different phase. For example, two or three rectangular functions are used to measure the distance in time-of-flight methods [7]. Well-designed base functions enable us to retrieve valuable information as an image from fewer data than the data needed by the event-driven approach. In the event-driven approach, every event has to be precisely detected and processed. Therefore, a huge amount of time-stamped data has to be processed on chip or read out and processed off chip. However, in our approach, the amount of information is significantly reduced by the charge-domain inner product calculation. In addition, due to the accumulation operation of modulated photoelectrons, the signal-to-noise ratio is increased effectively for image reproduction post-processing.

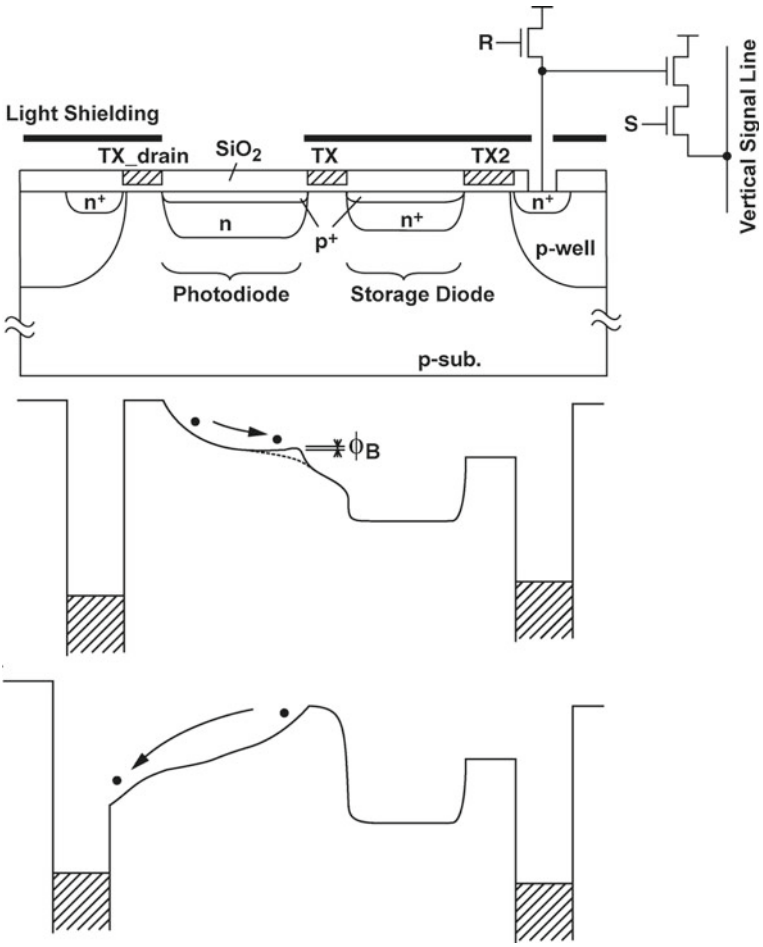


Fig. 6.7 Transfer-gate-based two-stage transfer (TST) pixel

6.3.2 Comparison with Transfer-Gate-Based Implementation

To clarify the advantages of the LEF control pixel, it is compared with a previous two-stage transfer (TST) pixel [8]. The TST pixel is more similar to the 4T pixel than is the LEF control pixel. Photoelectron transfer from the PPD to the PSD is controlled by a transfer gate, TX. Because the transfer gate is placed in the transfer path, barriers and dips can be generated around the transfer gate. The barrier ϕ_B in Fig. 6.7 disturbs fast electron transfer, and dips make the sensor insensitive in a very low light condition because the dips work as a charge trap. Although the transfer-gate-based charge transfer method is very common in CMOS image sensors, it is not suitable for fast and lossless photoelectron transfer.

6.3.3 Simple Implementation: Draining-Only Modulation

The draining-only charge modulation (DOM) pixel [9] shown in Fig. 6.8 is a simplified version of the LEF control pixel. G2 in Fig. 6.5 is omitted, and a fringing electric field is intrinsically prepared. In this pixel, a PPD is bridged to a PSD via a channel with a gate attached along the channel for draining all the photoelectrons. When the draining gate is closed, the photoelectrons generated and detected in the PPD are rapidly transferred to the PSD in less than a nanosecond if the channel is well engineered to prevent any potential barrier in the signal path from the PPD to the PSD. When the drain gate (G1) is opened, all the photoelectrons in the PPD are drained.

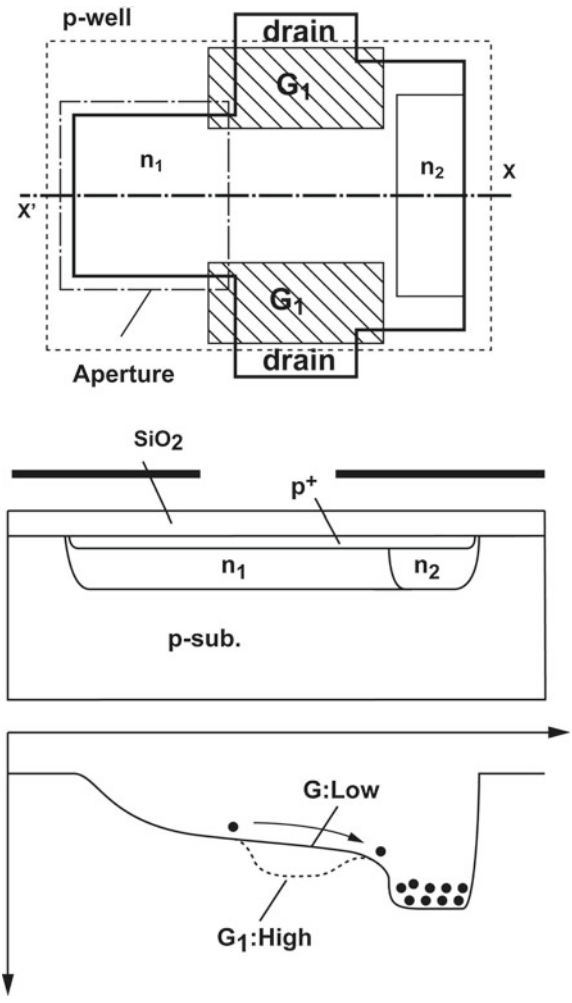


Fig. 6.8 Draining-only modulation pixel

6.4 Chip Implementation and Experiments

6.4.1 Fluorescence Lifetime Imaging Microscopy (FLIM) Image Sensor

One of the useful applications of DOM pixels is fluorescence lifetime imaging, which is becoming important in biological imaging. By using time windows controlled by the MOS gate, which have applied voltage represented by the TD in the DOM pixel, as shown in Fig. 6.9, the decaying process of the fluorescence can be observed with sub-nanosecond resolution. In the measurement, a single image is captured at a certain delay. A number of short light pulses are applied to a specimen, and the image is read out after accumulation of the modulated photoelectrons. To obtain the whole decay curve, the delay is scanned; that is, the time window slides as images are captured. Because the obtained images are an integral of the original decay curve inside a time window, the extinction parameter τ is determined by differentiating the series of captured images.

The structure of the DOM pixel and schematic potential diagrams are shown in Fig. 6.10. An intrinsic fringing electric field is formed by tapering the n-type region. The narrowest part is the PPD, whose potential is the shallowest, and the widest part is the PSD, whose potential is the deepest.

6.4.2 Experimental Results

A fabricated DOM image sensor and its specifications and measured results are shown in Fig. 6.11 and Table 6.1, respectively.

Figure 6.12 shows the measurement results of fluorescence decaying and the intrinsic response of the device. A time-resolved imager with DOM pixels

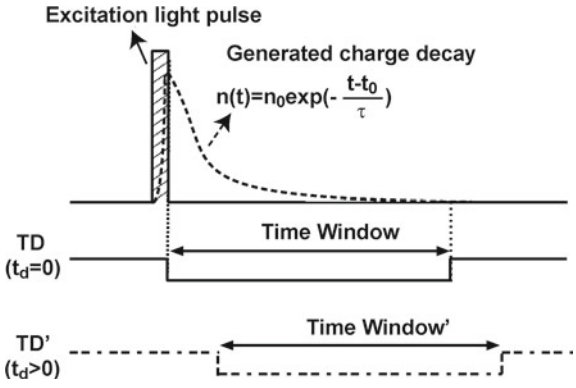


Fig. 6.9 Principle of decaying measurement by the DOM pixel

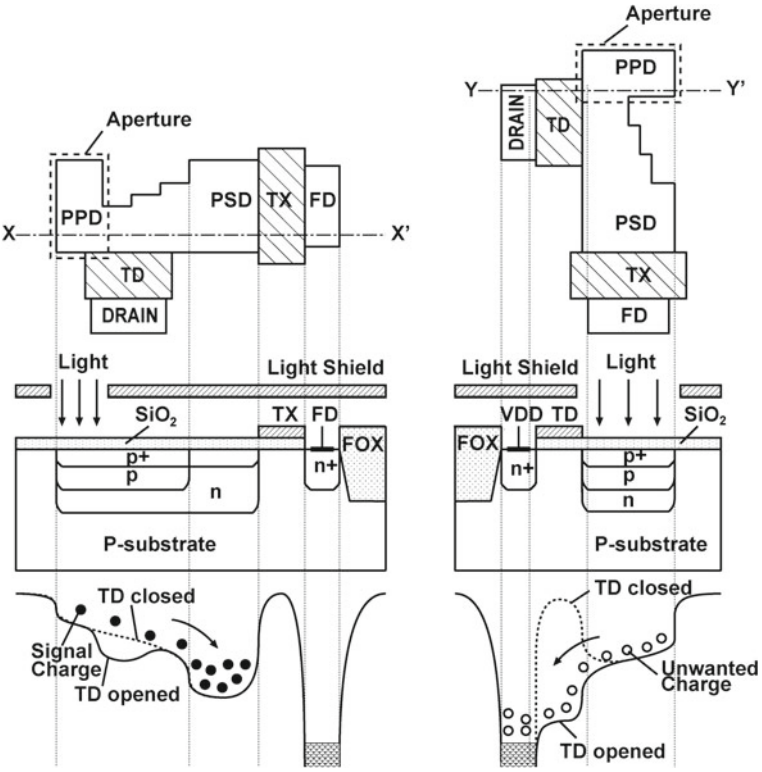


Fig. 6.10 Pixel structure and potential diagram of the DOM pixel

implemented in a 0.18-μm CMOS image sensor technology and a 374-nm semiconductor laser with a pulse width of 80 ps are used for the measurement [10]. In the present design, the measured lifetime of the DOM devices is 2 ns. If the fluorescence lifetime to be measured (target lifetime) is sufficiently higher than the device lifetime, the fluorescence lifetime is accurately measured. In Fig. 6.12, the measured lifetime of the green quantum dot having a target lifetime of 18.9 ns is 19.0 ns. If the target lifetime is comparable or even smaller than the device lifetime, the device lifetime affects the measured fluorescence lifetime. For example, a fluorescence lifetime having a target lifetime of 3 ns is measured to be 4.4 ns. This error can be reduced if a deconvolution technique using the impulse response of the DOM pixel is applied.

An important property of the DOM pixel is the single photoelectron sensitivity for time resolution on the nanosecond or sub-nanosecond time scale. As shown in Fig. 6.13, in the sensitivity measurement of the DOM pixel with a time window width of 125 ns, which is useful for measuring the fluorescence lifetime of nano- to a few tens of nanoseconds, the signal increase of the number of applied light pulses is linear at a very low light level of single photoelectron generation per 20 light pulses. This demonstrates that the DOM pixel can manipulate and detect a single photoelectron.

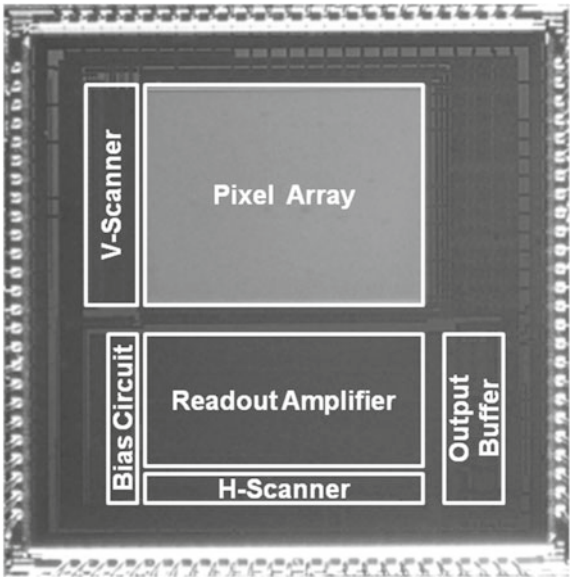


Fig. 6.11 Fabricated DOM image sensor

Table 6.1 Summary of specifications and measured results

Parameter	Value
Effective pixel	256(Column) × 256(Row)
Pixel size	7.5 μm × 7.5 μm
Frame rate	15 fps
Fill factor (without microlens)	4.6 %
Conversion gain	30 μV/e$^-$
PSD full well capacity	3800 e$^-$
Extinction ratio	30:1 (@ 440 nm)
Dark signal generation ratio	1e$^-$ per 546 TD pulses
Temporal noise @ 300 K	17.2e^-_{rms} (@ *CDS Gain* 1.07×)
	2.0e^-_{rms} (@ *CDS Gain* 30×)
Intrinsic lifetime	2 ns

To further demonstrate fluorescence decay imaging, the intensity decay images of HeLa cells stained with two types of fluorophores are shown in Fig. 6.14. The cell nucleus is stained by a Hoechst dye with a measured lifetime of 5.6 ns, and the cytoplasm is stained by quantum dots (QD) with a measured lifetime of 14.0 ns.

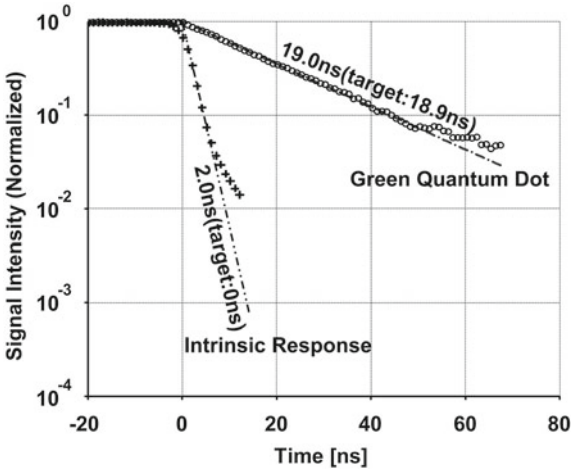


Fig. 6.12 Measured decaying responses

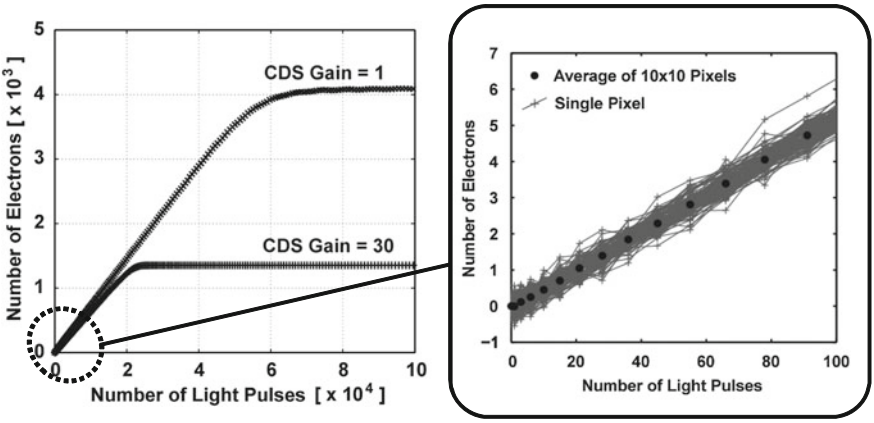


Fig. 6.13 Relationships between light intensity (number of light pulses) and detected electrons

The delay time is shown at the bottom of each image. This figure demonstrates that the fluorescence by Hoechst dye decays more quickly than that by the quantum dots.

6.5 Conclusions

In this chapter, two-dimensional lateral electric field (LEF) control for single photoelectron manipulation and detection with sub-nanosecond resolution has been described. The draining-only modulation (DOM) pixel as a time-resolved single

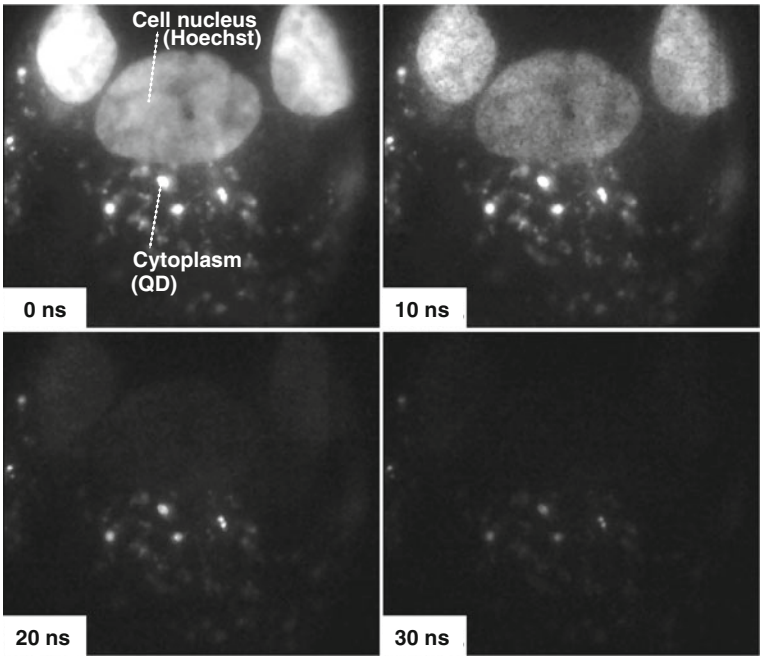


Fig. 6.14 Captured images of fluorescence decays of HeLa cells

photoelectron detector was introduced and its application to fluorescence lifetime imaging was described as an example of single photoelectron sensitivity in highly time-resolved imaging. The photoelectron modulation technique using two-dimensional or three-dimensional electric field control is expected to be a powerful tool for future applications in biological imaging, such as fluorescence correlation spectroscopy and Raman spectroscopy, and in industrial imaging, such as very high resolution time-of-flight range measurement.

Acknowledgments This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology under a Grant-in-Aid for Scientific Research (A), No. 22246049. The authors are grateful to Prof. Hashimoto, Prof. Niioka, Mr. Ukon, Dr. Li, Ms. Baek, and Mr. Han.

References

1. A. Periasamy, R. Clegg, *FLIM Microscopy in Biology and Medicine* (CRC Press, Boca Raton, 2009)
2. C. Freudiger, W. Min, B. Saar, S. Lu, G. Holton, C. He, J. Tsai, J. Kang, X. Xie, Label-free biomedical imaging with high sensitivity by stimulated Raman scattering microscopy. Science **322**, 1857–1861 (2008)

3. C. Niclass, A. Rochas, P.-A. Besse, E. Charbon,Toward a 3-D camera based on single photon avalanche diodes. IEEE J. Sel. T. Quant. Elec. **10**(4), 796–802 (2004)
4. T. Spirig, P. Seitz, O. Vietze, F. Heiger, The lock-in CCD-two-dimensional synchronous detection of light. IEEE. J. Quant. Electron. **31**(9), 1705–1708 (1995)
5. N. Teranishi, A. Kohno, Y. Ishihara, K. Arai, No image lag photodiode structure in the interline CCD image sensor. in *IEDM Technical Digest* (1982) pp. 324–327
6. M.-W. Seo, S.-H. Suh, T. Iida, T. Takasawa, K. Isobe, T. Watanabe, S. Itoh, K. Yasutomi, S. Kawahito, A low-noise high intrascene dynamic range CMOS image sensor with a 13 to 19b variable-resolution column-parallel follding-integration/cyclic ADC. IEEE J. Solid-State Circ. **47**(1), 272–283 (2012)
7. S. Kawahito, I.A. Halin, T. Ushinaga, T. Sawada, M. Homma, Y. Maeda, A CMOS time-of-flight range image sensor with gates-on-field-oxide structure. IEEE Sens. J. **7**(12), 1578–1586 (2007)
8. H.J. Yoon, S. Itoh, S. Kawahito, A CMOS image sensor with in-pixel two-stage charge transfer for fluorescence lifetime imaging. IEEE Trans. Electron Dev. **56**, 214–221 (2009)
9. S. Kawahito, Z. Li, K. Yasutomi, A CMOS image sensor with draining only demodulation pixels for time-resolved imaging, in *Proceedings of 2011 International Image Sensor Workshop*, Hakodate (2011) pp. 185–188
10. Z. Li, S. Kawahito, K. Yasutomi, K. Kagawa, J. Ukon, M. Hashimoto, H. Niioka, A time-resolved CMOS image sensor with draining only modulation pixels for fluorescence lifetime imaging. IEEE Trans. Electron Dev. **59**(10), 2715–2722 (2012)

Chapter 7
Engineering of a Nanometric Optical System Based on Optical Near-Field Interactions for Macro-Scale Applications

Naoya Tate, Makoto Naruse and Motoichi Ohtsu

Abstract The operating principle of our nanometric optical system is based on its hierarchical structure. This hierarchy is due to the characteristic behavior of optical near-fields induced by nanometric interactions between light and materials. In order to exploit such nanometric phenomena in macro-scale applications, it is essential to design and engineer each component of the system so that the optical near-field behavior can be extracted as macro-scale optical responses. Here we introduce the basic concept of our nanometric optical system, and we describe some demonstrations of macro-scale applications whose operation is based on this concept.

7.1 Introduction

Conventional optical phenomena are described by various physical quantities of light, such as intensity, wavelength, phase, and polarization. In the research field of optical computing [1] and its related technologies, several characteristics of conventional *propagating* light are used for implementing multidimensional information elements to realize novel kinds of information processing. The propagation speed, spatial parallelism, and multi-dimensionality of related phenomena made significant contributions to a number of developments in the 1980s, particularly the progress made in high-speed optical communications and high-density optical storage. On the other hand, conventional propagating light exhibits wave-like behavior

N. Tate (✉) · M. Ohtsu
The University of Tokyo, Tokyo, Japan
e-mail: tate@nanophotonics.t.u-tokyo.ac.jp

M. Ohtsu
e-mail: ohtsu@ee.t.u-tokyo.ac.jp

M. Naruse
National Institute of Information and Communications Technology, Tokyo, Japan
e-mail: naruse@nict.go.jp

M. Naruse (ed.), *Nanophotonic Information Physics*,
Nano-Optics and Nanophotonics, DOI: 10.1007/978-3-642-40224-1_7,

in macro-scale space, and the diffraction limit has recently become a high barrier to further advances demanded by our growing information society, including not only faster processing and more highly integrated systems, but also energy and resource saving in the construction and operation of information systems.

We have been studying applications that utilize the characteristic behavior of *dressed photons* [2], which are often described as *non-propagating* light induced by irradiation of a nanometric component with propagating light. Because the wavelength of a dressed photon cannot be physically defined, the diffraction limit does not come into play. This has allowed us to realize both *quantitative* innovations, where the applicable scales are much smaller than the diffraction limit of light, and *qualitative* innovations, which reveal novel optical functions that are uniquely realized only by using dressed photon technologies.

Our nanometric optical system is one example exploiting such innovations. The fundamental principle of our nanometric optical system is based on its intrinsic *hierarchical structure* [3], which allows multi-step interfacing from nano-scale to macro-scale space due to *optical near-field interactions* induced between nanometric components of the system. By appropriately designing and systematically engineering nanometric components, we experimentally verified that the characteristic behaviors of optical near-fields in nano-scale space can be successfully used as novel optical functions in macro-scale space.

In this Chapter, first we describe the basics of nanophotonics based on the concept of dressed photons. Then, we describe the concept of our nanometric optical system and introduce some demonstrations of actual nanometric optical systems.

7.2 Nanometric Optical System

7.2.1 Nanophotonics

Nanophotonics is a new optical technology utilizing local interactions between nanometric components via optical near-fields [2]. The optical near-fields are schematically described as a virtual cloud of photons that is constantly localized around nanometric components illuminated by incident light, as shown in the left hand side of Fig. 7.1. In this situation, a real photon, which corresponds to scattered light, can be emitted from an electron in an illuminated nanometric component. As shown in the right hand side of Fig. 7.1, represented by a Feynman diagram, independently of the real photon, another photon is emitted from the electron, and this photon can be re-absorbed within a short time. This photon, which is regarded as a virtual photon, is nothing more than the optical near-field, and its energy is localized at the surface of the nanometric component. Since the virtual photon remains close to the electron, it can couple with the electron in a unique manner. This coupled state, called a *dressed photon*, is regarded as a quasi-particle that contributes to exciton energy transfer [4–6].

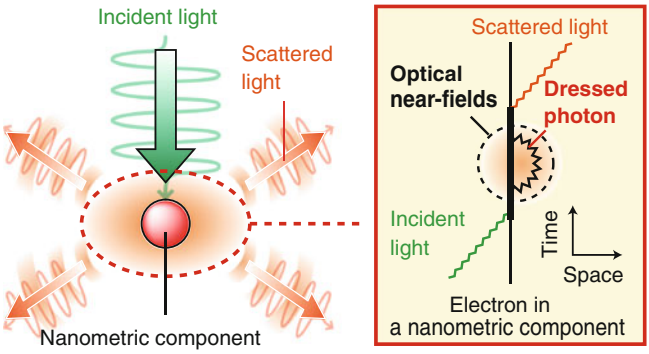


Fig. 7.1 Schematic diagram of induced optical near-field (*left*), and description of optical near-field and dressed-photon using Feynman diagram (*right*)

Various characteristic features of dressed photons have been actively studied from both theoretical and experimental aspects [2]. As a result, dressed photons have allowed the realization of a number of qualitative innovations in optical technologies that would not be possible if only conventional propagating light were used. Examples include novel optical devices [7–12], fabrication technologies [13–15], energy conversion [16, 17], and information systems [3, 18, 19]. The achievements of dressed photon technology are also expected to offer advantages in energy saving and environmental protection for future society.

7.2.2 Basics of Nanometric Optical System

The spatial distribution of an optical near-field defines the scale of the interactions between multiple nanometric components, which are represented by the exchange of dressed photons between them. The optical response of a single nanometric electron system, which consists of the minimum number of components exhibiting the required functions, as shown in Fig. 7.2a, can be extracted to macro-scale space by a chain of elemental steps from a single setup to a more complex setup consisting of a number of integrated systems. We define such a system structure as a *nanometric optical system*. As conceptually shown in Fig. 7.2b, the most important characteristic of a nanometric optical system is its *hierarchical structure* [3], which consists of multiple accessible layers in an information space extending from the nano-scale to the macro-scale. In such a structure, two types of electronic actions are induced: an action terminating in each single layer, and an action connecting to a larger-scale layer. Moreover, the most important point is that such actions are induced independently and can occur individually. Exploiting this enables us to realize novel information systems in which a large number of functions are highly integrated in a single system by appropriately associating each function with the corresponding action revealed at each layer of the system.

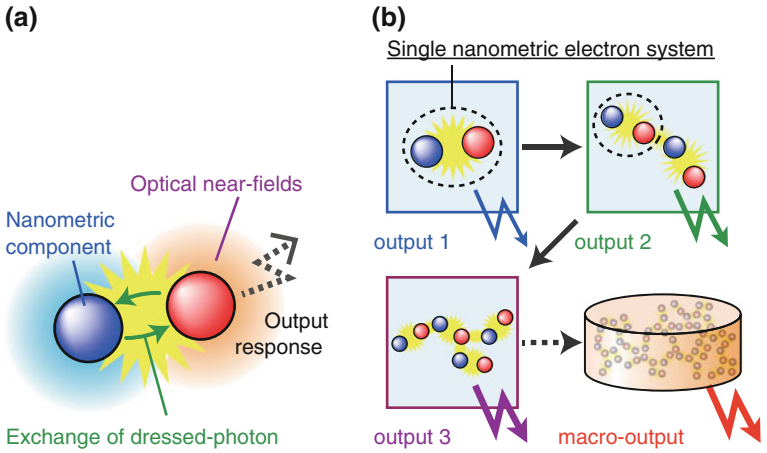


Fig. 7.2 Schematic diagram of **a** a single nanometric electron system and **b** concept of hierarchical structure of nanometric optical system due to a chain of optical near-field interactions between nanometric electron systems

The behavior of each nanometric electron system depends on the physical properties, shape, and size of its components. On the other hand, the chain of elemental steps depends on the spatial alignment of the nanometric electron systems and their interaction with the environmental system. What is fundamental in constructing a nanometric optical system is appropriate design and engineering of not only each single nanometric electron system but also the environmental system and the interaction mechanism between them.

This hierarchy of the nanometric optical system is important when discussing interfacing from the scale of optical near-fields to that of optical far-fields. Namely, by accessing an arbitrary layer in the system, we can extract the behavior of optical near-field interactions in the form of an observable optical far-field response at the macro-scale, which is a result of the chained elemental steps. As demonstrations of this concept, we previously reported a method for magnified transcription of an optical near-field distribution by using a photo-induced phase transition in a metal complex [20], and polarization conversion utilizing a quadrupole–dipole transformation based on optical near-field interactions between precisely aligned nanometric metal structures [21, 22]. Both of these ideas were realized by defining appropriate correspondences between the nanometric behavior of electrons in nano-scale space and the optical responses in macro-scale space, in order to directly retrieve the nanometric behavior in the form of macro-scale optical responses. Moreover, a hierarchical nanophotonic hologram [19, 23, 24] has been successfully developed as a novel optical security device in which both covert and overt security are revealed in nano-scale space and macro-scale space, respectively.

7.3 Modulatable Nanophotonics

7.3.1 Concept

A semiconductor quantum dot (QD) is a three-dimensional nanometric structure that can confine an injected electron, hole, and related exciton. It has quantum energy levels due to the quantum size effect. Because of their small size and stable absorption and emission properties, QDs have been used as a source of optical near-fields and as basic components of nanophotonic devices which exhibit functional optical responses by utilizing optical energy transfer between QDs via the induced optical near-fields. Recently, various experimental demonstrations, such as logic gates, transmission channels, and optical switches, have been achieved as typical examples of nanophotonic devices [7–12].

The operation of the nanophotonic devices described above exhibits a one-to-one correspondence with respect to input signals, because the physical properties, size, shape, and alignment of the components for the characteristic optical near-field interactions are fixed on each substrate. In order to realize a one-to-many correspondence with a single nanophotonic device, it is necessary to implement modulatable optical near-field interactions and corresponding optical far-field functions. Here, we describe the novel concept of *modulatable nanophotonics* to realize such a system [25] . As schematically shown in Fig. 7.3, it is realized by providing appropriate external controls to modulate the parameters of the components. In our concept, optical far-field retrieval of induced optical near-field interactions is achieved by modulating the intensity, the polarization, and the spectra of the subsequent emissions from QDs. This is one of the most characteristic features of our nanometric optical system; that is, the effect of nano-scale interactions can be directly revealed as macro-scale optical outputs by appropriately designing and engineering the setup.

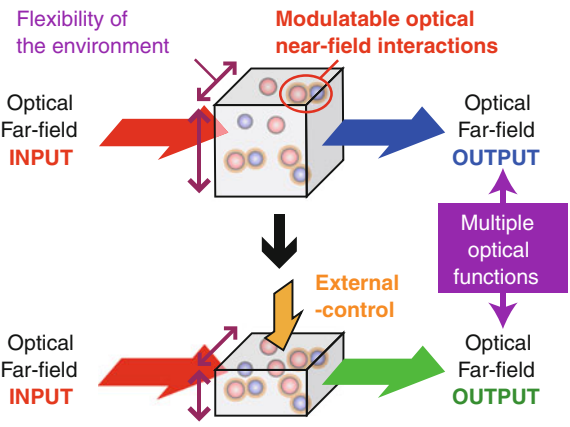


Fig. 7.3 Conceptual diagram of one-to-many correspondence in nanophotonic device based on concept of modulatable nanophotonics

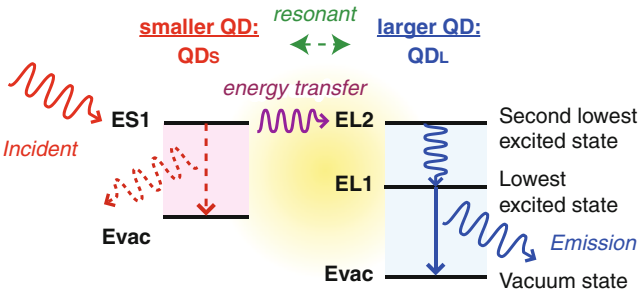


Fig. 7.4 Schematic diagram of local energy transfer between closely positioned smaller and larger QDs. Energy transfer and subsequent energy dissipation are allowed only via optical near-fields induced between the two QDs

7.3.2 Basics of Optical Energy Transfer

Here we describe the basics of energy transfer between size-selected QDs via induced optical near-field interactions. Figure 7.4 schematically shows smaller and larger quantum dot structures QD_S and QD_L, respectively. The ratio of their sizes is strictly controlled to be $1{:}\sqrt{2}$ in the case of a cubic structure, so that they both satisfy the energy condition that the ground level E_{S1} of QD_S and the first excited level E_{L2} of QD_L are resonant with each other. Because the wave function of E_{L2} in QD_L is an odd function, E_{L2} is generally assumed to be an optically forbidden level that cannot be excited by propagating light. However, in the case where the two QDs are closely positioned, excitation of QD_L is allowed via local electrical divergence due to the optical near-fields induced at the neighboring QD_S. As a result, energy transfer from E_{S1} to E_{L2} is allowed via excitation of the barrier layer in both QDs. The transferred energy is immediately dissipated from E_{L2} to the lowest excited state E_{L1} of QD_L, and a photon is subsequently emitted. Experimental and theoretical results showing good agreement with each other have been discussed in detail in previous papers [5, 26–29]. On the other hand, emission from QD_S cannot be observed.

In the case where the two QDs are sufficiently separated from each other, no interactions are induced between the two QDs, and the two QDs emit photons individually. The required size of this separation is defined by the scale of the optical near-field interactions and corresponding size-resonant effects [30]. Because the induced optical near-fields are localized around nanometric structures, and the scales of their spatial distributions are comparable to the sizes of the structures, as described by a Yukawa function [31], the energy of the optical near-field interactions between two nanometric structures is expected to be maximized when the sizes of the two structures are similar. This indicates that the separation limit for the optical energy transfer via the optical near-field interactions is similar to the size of each QD.

7.3.3 Modulated Emission Based on Optical Energy Transfer

Here, we describe the basic concept of modulatable nanophotonics, in which optical near-field interactions and the corresponding optical far-field responses are modulated. The concept can be experimentally verified by utilizing energy transfer between closely positioned resonant QD pairs dispersed on a flexible substrate. Modulation by physical flexion of the substrate changes the distances between the QDs to control the amount of optical energy transfer between neighboring QDs.

As described in Sect. 7.3.2 above, closely positioned resonant QD pairs exhibit high-quantum-efficiency and selective energy transfer via induced optical near-fields between the QDs. Here, we consider the case where QDs composed of the same material are dispersed on a substrate having sufficient physical flexibility. When the substrate is flexed, the distance between the QDs can be varied. Figure 7.5a, b show schematic diagrams comparing the emission processes in resonant and non-resonant QD pairs, respectively. In the case of a resonant QDs pair, denoted QD_{S-R} and QD_{L-R} in Fig. 7.5a, energy transfer occurs when the QDs are sufficiently close, because the magnitude of the optical near-field interaction, i.e., the coupling strength between QD_{S-R} and QD_{L-R}, depends on the distance r between them, as represented by the Yukawa function:

$$U = \frac{A\exp(-\mu r)}{r}, \tag{7.1}$$

where μ and A represent the distribution of optical near-fields, which are determined by the exciton energy and the dipole moment, respectively [31]. Thus, emission is preferentially obtained from the lowest excited state of QD_{L-R}. In contrast, if the separation between QD_{S-R} and QD_{L-R} is increased by flexing the substrate in order to significantly decrease the coupling strength, both QD_{S-R} and QD_{L-R} emit light independently. This means that the spectral intensities of QD_{S-R} and QD_{L-R}, as well as the relative spectral intensity ratio between them, can be modulated by flexing, i.e., by modulating, the coupling strength. Thus, the spectral intensity ratios between QD_{S-R} and QD_{L-R} in the cases with and without flexion are evidently different. On the other hand, in the case of non-resonant QDs, denoted QD_{S-NR} and QD_{L-NR} in Fig. 7.5b, energy transfer never occurs between them. Each QD emits light individually, regardless of whether the substrate is flexed. In this case, only a change in spectral intensity that depends on the areal density of the QDs is obtained. Therefore, the spectral intensity ratios between QD_{S-NR} and QD_{L-NR} in the cases with and without flexion are always equal.

Here, we define a figure-of-merit called the *modulatability*, M_{sp}, of the emission by using the spectral intensities $I(e_S)$ and $I(e_L)$, which correspond to the peak optical energies in the emission spectra of QD_S and QD_L, respectively:

$$M_{sp} = \left| \frac{I'(e_S)}{I(e_S)} - \frac{I'(e_L)}{I(e_L)} \right|, \tag{7.2}$$

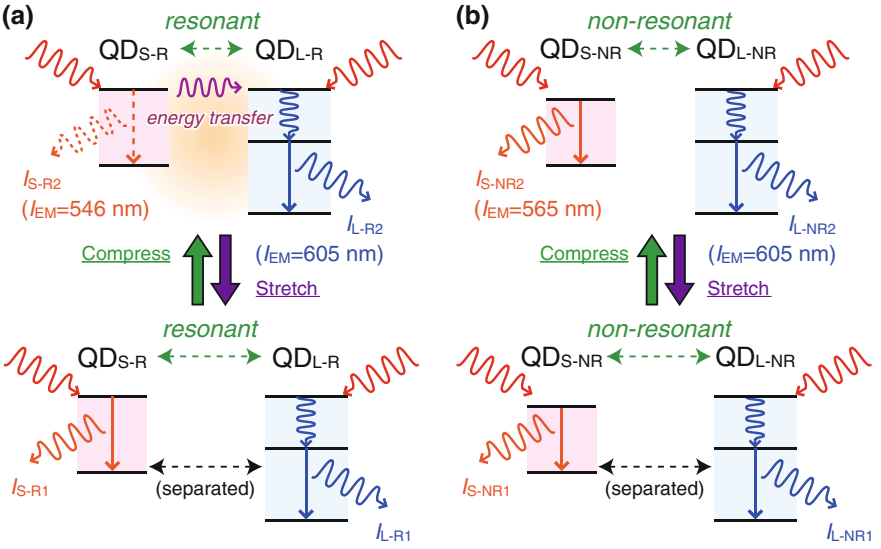


Fig. 7.5 Schematic diagram of **a** modulatable optical energy transfer between resonant QD pairs dispersed on a flexible substrate and **b** non-modulatable, non-resonant QD pairs

where I and I' represent the peak emission intensities of the spectra without and with flexion, respectively. In the case of the resonant QDs, the value of I'/I depends on the magnitude of the coupling strength, which is governed by the flexion, which means that M_{sp} is calculated to be a non-zero value ($M_{\mathrm{sp}} \neq 0$). On the other hand, in the case of the non-resonant QDs, the value of I'/I remains unchanged for all wavelengths because the areal density of the QDs is homogeneously modulated by the flexion. Therefore, M_{sp} is calculated to be nearly zero ($M_{\mathrm{sp}} \simeq 0$). One remark regarding 7.2 is that singularity may be a problem when the denominators are zero. However, we can assume that the denominators are non-zero in realistic physical situations, which guarantees the validity of the representation in 7.2.

7.3.4 Numerical Demonstration

To numerically estimate the spectral intensity modulation based on the proposed concept, we assumed a calculation model consisting of four QDs configured as resonant QD pairs, i.e., two small QDs ($\mathrm{QD}_{\mathrm{S-A}}$ and $\mathrm{QD}_{\mathrm{S-C}}$) and two large QDs ($\mathrm{QD}_{\mathrm{L-B}}$ and $\mathrm{QD}_{\mathrm{L-D}}$), representing many QDs dispersed on a substrate, as shown in Fig. 7.6a. We simulated temporal evolution of the exciton populations of the relevant excited states in these QDs using quantum master equations [32]. The exciton populations correspond to the intensity of radiation from each QD, and their evolution is modulated by modulating the magnitude of the optical near-field coupling

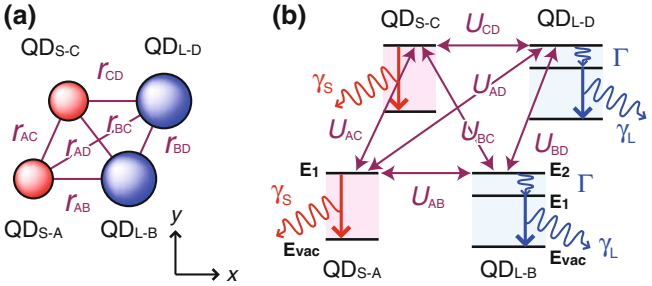


Fig. 7.6 Schematic diagrams of four-QD calculation model for numerical demonstration of multiple-QD system, showing **a** their geometrical alignment and **b** the conditions of optical near-field interactions between each QD. In our model, each parameter is variable, and their variations are assumed to be imposed by flexion of the substrate

strength between QDs, which corresponds to flexion of the substrate. As previously reported in [33], the initial condition is set as a vacuum state, and the duration and the amplitude of the incident light are given by a Hamiltonian representing interactions between the incident light and the QDs. The magnitude of the optical near-field interaction, i.e., the coupling strength, between the two QDs is denoted by the Yukawa function of 7.1.

In this model, the evolution of the exciton population in each QD is calculated for various distances r, which correspond to different degrees of coupling, controlled by flexion. The initial distances between each QD on a 2D xy-surface are set to $r_{AB} = r_{BD} = r_{AC} = r_{CD} = 100$ (nm), and the magnitudes of the coupling strengths U_{AB}, U_{BD}, U_{AC}, U_{CD}, U_{AD}, and U_{BC} depend on these distances, as schematically shown by Fig. 7.6b. In our model, the relaxation time constant Γ^{-1} is set at 10 ps, and the radiation lifetimes of QD_S and QD_L, denoted as γ_S^{-1} and γ_L^{-1}, are respectively set at 2.8 and 1.0 ns, according to a typical parameter set for CdSe-QDs used in our previous experiments [9], which is based on the pattern of conventional energy transfer models between CdSe-QDs. Based on these parameters and the proportional relation $A \propto \mu \propto L^3$ in 7.1, the diameter of the smaller QDs, L, is assumed to be approximately 50 nm, and that of the larger QDs, L', satisfies $L' = \sqrt{2}L$. Because the scale of optical near-fields is inversely proportional to the square of the distance from the source, sufficient energy transfer between QDs via optical near-fields is expected with such a model.

First, we assumed stretching of the substrate in the x and y directions, as shown in Fig. 7.7a, and calculated the exciton populations with various stretch lengths δr. In the case of the x-directional stretching, r_{AC} and r_{BD} are constant, in a first-order approximation, and r_{AB}, r_{CD}, r_{AD}, and r_{BC} are varied. On the other hand, in the case of the y-directional stretching, r_{AB} and r_{CD} are constant, and r_{AC}, r_{BD}, r_{AD} and r_{BC} are varied. Figure 7.7b shows the calculated results. The horizontal axis represents the relative stretch length, which is defined as $\delta r/L$.

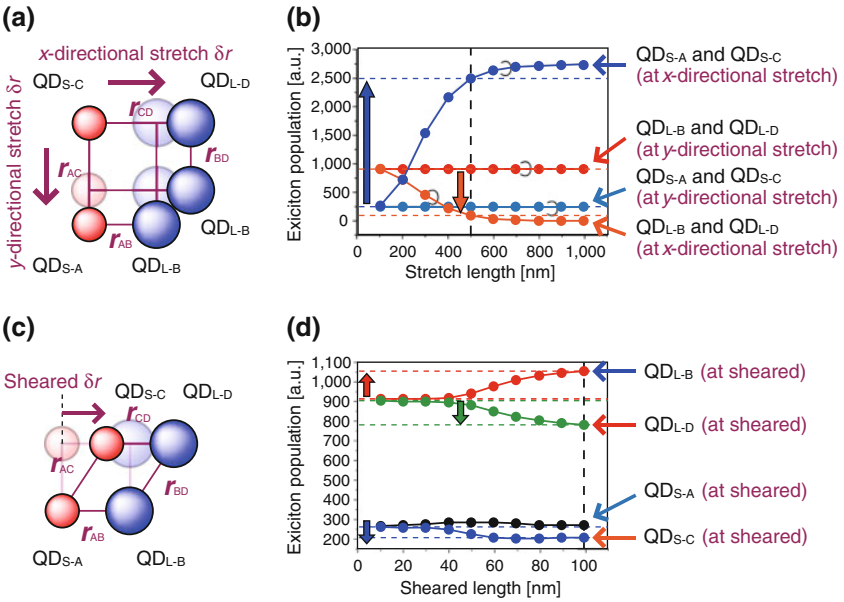


Fig. 7.7 Schematic diagram of **a** stretch model and **c** shear model, and evolutions of the exciton populations (emission intensity) from four QDs by **b** stretching and **d** shearing the substrate

Evident changes in the exciton populations are obtained by the x-directional stretching, because the excitons always preferentially transfer from smaller QDs to larger QDs (from QD_{S-A} to QD_{L-D}). This result indicates that the coupling strength between the smaller QDs and the larger QDs is decreased by the x-directional stretching, and they emit independently.

Here we assume that the emission intensity from each QD is proportional to the exciton population at the corresponding QD energy level. Therefore, in the case of the x-directional stretching, M_{sp} is calculated to be 9.26 at a stretch length of $\delta r = 500\,\text{nm}$. The population in the lowest excited state of the larger QDs finally approaches zero, and M_{sp} is calculated to be infinity. On the other hand, y-directional stretching gives $M_{sp} \simeq 0$ at the same stretch length. Therefore, only the former model can be said to be a *modulatable nanophotonic system*.

Next, a shear model is assumed for the x-directional shift of QD_{S-C} and QD_{L-D}. A schematic diagram is shown in Fig. 7.7c. In the case of the model, r_{AB} and r_{CD} are constant, and r_{AC} and r_{BD} are increased. Figure 7.7d shows the results of calculations. By increasing the shear length δr, which is defined by the shifted values of QD_{S-C} and QD_{L-D}, only the exciton population at QD_{L-B} increases, because the magnitude of energy transfer toward QD_{L-B} increases not only from QD_{S-A} but also from QD_{S-C}. In contrast, that to QD_{L-D} decreases. This is also because of the increase in the energy transferred from QD_{S-C} not only to QD_{L-D} but also to QD_{L-B}. As a result, the population at QD_{S-C} is decreased, whereas that at QD_{S-A} remains

unchanged, because the main route of energy transfer is from QD_{S-A} to QD_{L-B}, regardless of the stretching. By these contributions of the exciton populations at the smaller QDs (QD_{S-A} and QD_{S-C}) and the larger QDs (QD_{L-B} and QD_{L-D}), the calculated M_{sp} in the shear model shows a non-zero value ($M_{sp} = 0.11$ for a shear length of $\delta r = 100\,\text{nm}$).

7.3.5 Experimental Demonstration

We conducted an experiment to observe the change in energy transfer between QDs and related emission spectra before and after flexion of each substrate. We used commercially available CdSe-QDs (Ocean Optics Inc.; *Evidot*) as test specimens. The QDs were uniformly dispersed in a toluene solvent at a concentration of 10 mg/mL. Their exciton energy transfer has already been studied [4–6], and resonant and non-resonant conditions have been experimentally verified [9]. The peak absorption optical energy e_{AB} and peak emission optical energy e_{EM} of large QDs, QD_{L-R} and QD_{L-NR}, were as given in the technical specifications provided by the manufacturer, namely, 2.13 and 2.05 eV, respectively. The QDs used as QD_{S-R} had $e_{AB} = 2.37\,\text{eV}$ and $e_{EM} = 2.27\,\text{eV}$, and those used as QD_{S-NR} had $e_{AB} = 2.19\,\text{eV}$ and $e_{EM} = 2.15\,\text{eV}$. This emission optical energy e_{EM} corresponds to the optical energy of emission from the lowest excited state of each QD. The respective diameters D of the resonant pair, QD_{S-R} and QD_{L-R}, were assumed to be 8.2 and 8.7 nm, and those of the non-resonant pair, QD_{S-NR} and QD_{L-NR}, were assumed to be 7.7 and 8.7 nm. For the experimental demonstration, we used polydimethylsiloxane (PDMS), which is particularly known for its strong rheological properties. We mixed 5 mL of each QD solution to prepare the resonant and non-resonant QD pairs. The mixed QD solutions were uniformly dispersed on a 2 cm $\times$ 2 cm square of PDMS substrate and allowed to dry naturally. In our experiments, although evident heterogeneity of the distributions was observed, the average r was assumed to be approximately $5 \sim 10\,\text{nm}$ from the thickness of the shell and the length of the modified ligand to each QD. The QDs were excited by a He-Cd laser (photon energy 3.82 eV) with a power density of $5\,\text{mW/cm}^2$. In our experiment, the PDMS substrate was set on an aperture formed at the side of a vacuum desiccator and was flexed by evacuation. The flexion brought the dispersed QDs closer to each other, as represented by $\delta r < 0$. The air pressure was decreased and fixed to $\sim$0.07 MPa to achieve a 20 % in-plane compression ratio of the substrate, which was geometrically determined from the size of the aperture and the depth of the flexed substrate.

We experimentally observed the emission spectra of the resonant and non-resonant QDs without and with flexion by using a spectrometer (JASCO; CT-25TP), as shown in Fig. 7.8a, b, respectively. The fractional ratio of the numbers of small QDs to large QDs was 1:1. The smoothed curves in these figures represent Gaussian curves fitted to the measured spectral profiles. The differences in peak optical energy from those shown in the technical specifications were due to the variability of each particle's size and the inhomogeneous dispersion of the size distribution.

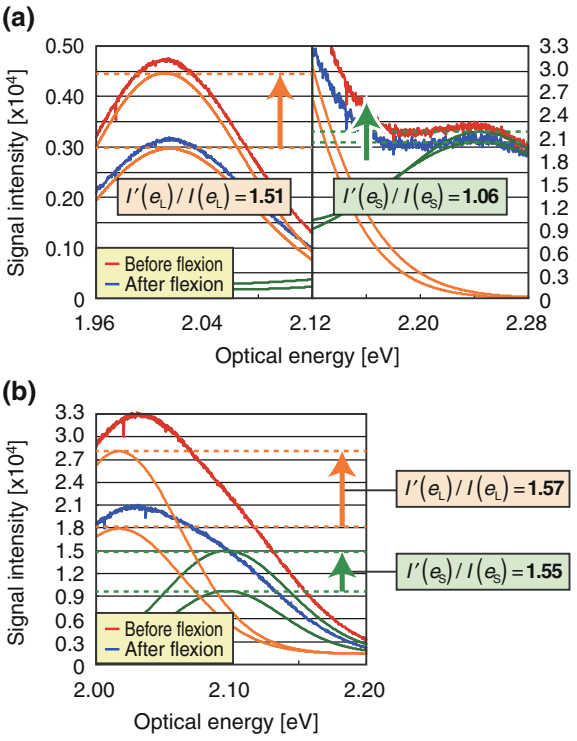


Fig. 7.8 Obtained emission spectra with **a** resonant and **b** non-resonant samples before (*blue curves*) and after (*red curves*) flexion of the substrate. Dashed curves represent the results of Gaussian fitting of each spectrum

As shown, the intensities of the spectra were increased as a whole because the number of QDs per unit area was increased by flexing each substrate. However, only an increase of the emission intensity from QD_{S-R} was suppressed in the case where the resonant pair was used. This is because the average distance between QD_{S-R} and QD_{L-R} is shortened by the flexion, which induces energy transfer from QD_{S-R} to QD_{L-R}. Similar behavior was predicted by our numerical demonstrations described above. To quantitatively evaluate this behavior, we examined the intensities at the spectral peak from the fitted curves and calculated the values of $I'(e_L)/I(e_L)$ and $I'(e_S)/I(e_S)$, as shown in Fig. 7.8a, b. By taking the difference, the modulatabilities M_{sp} obtained with the resonant and non-resonant QDs were 0.45 and 0.02, respectively. Therefore, as in the numerical demonstration in Sect. 7.3.4, it was confirmed that the resonant QDs showed a much larger M_{sp} compared with that of the non-resonant QDs.

The difference in the rates of change of each spectral peak intensity for the resonant QD sample corresponds to a difference in color tone of the emission from the sample. Our experimental results indicated that a change in the emitted color tone was observed by flexion of the substrate on which the resonant QDs were dispersed. We constructed a chromaticity diagram based on the color matching functions of each observed emission spectrum and defined another modulatability M_{ch} to directly

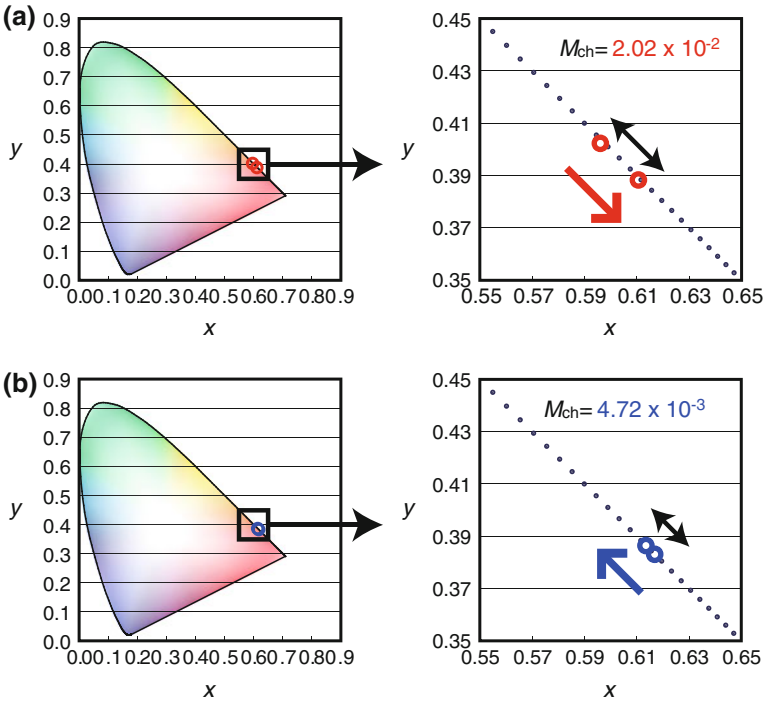


Fig. 7.9 Calculated modulatabilites M_{ch} with **a** resonant and **b** non-resonant samples. An evident difference in M_{ch}, defined as the distance in the chromaticity coordinates before and after flexion, was observed

evaluate the amount of each modulation for several samples. The value of M_{ch} is defined as the shift between coordinates of each spectrum in the chromaticity diagram due to the modulation. Figure 7.9a, b show chromaticity coordinates of previous experimental results and corresponding values of M_{ch}. As shown, the resonant QD sample exhibited larger values of M_{ch} (Fig. 7.9a), whereas the non-resonant QD sample exhibited relatively small values of M_{ch} (Fig. 7.9b). Because we used the same materials for the QD pairs in our experiment, similar shift vectors from the two coordinates on the chromaticity diagram were obtained. That is to say, we can evaluate each modulation, not only in terms of the amount but also the material, based on possible directions of each vector from before modulation to after modulation. Variation of the material is another fundamental issue that must be considered in applications utilizing one-to-many correspondences.

7.3.6 Conclusion

We have described the basic concept of modulatable nanophotonics as a typical demonstration of the interfacing between nano- and macro-scale spaces. Numerical and experimental demonstrations of the concept utilizing QDs dispersed on flexible substrates have been described. The modulatability was qualitatively evaluated by introducing the figures-of-merit M_{sp} and M_{ch}, and resonant QD pairs exhibited unique modulatability of their emission intensities. Moreover, such modulatability was not observed by using non-resonant QD pairs. As we numerically and experimentally verified, the modulatability fundamentally depends on the amounts of energy transfer due to the coupling strength between the QDs via induced optical near-fields. From the viewpoint of the conceptual nanometric optical system described here, the results of our demonstration can be regarded as an actual example of optical interfacing between the nano-scale and the macro-scale. The evident modulation of the emission spectra at the macro-scale is fundamentally due to nano-scale changes in the distances between QDs.

7.4 Nanophotonic Droplet

7.4.1 Concept

The time and energy resources required for local electromagnetic interactions between nanometric components via optical near-fields are very small. The high-speed performance, energy efficiency, and compactness of several nanophotonic devices have been experimentally demonstrated [7, 9, 12, 34, 35]. Such a discussion can be similarly applied to our nanometric optical system as well. In order to realize further improvements for future practical use, several autonomous techniques for arranging QDs and metal materials in the intended structures have been developed [14, 36–38].

Here we describe a novel method for selective and autonomous pairing of particular QDs by utilizing a phonon-assisted optical near-field process [16]. Paired QDs are subsequently encapsulated into a submillimeter-scale structure forming a *nanophotonic droplet* [39], and they are appropriately aligned to induce optical near-field interactions. Therefore, as conceptually illustrated in Fig. 7.10, although each QD serving as a component of a nanophotonic droplet exhibits its intrinsic behavior in nano-scale space, the nanophotonic droplet as a whole exhibits high-quantum-efficiency optical-near-field interactions and corresponding optical far-field responses, so that it can be directly utilized in various macro-scale applications.

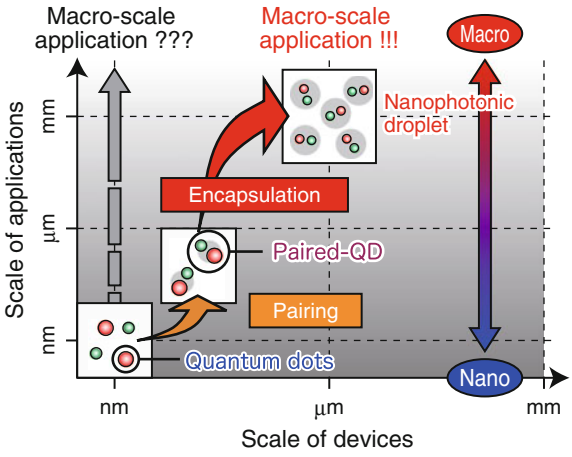


Fig. 7.10 Conceptual diagram of integration and utilization of nanophotonic droplets for macro-scale applications

7.4.2 Basics

We have demonstrated that a dressed photon can excite a multi-mode phonon in a nanometric material, and that excited phonon states can couple with each other [17, 36, 38, 40]. The quasi-particle representing this coupled state has been named a dressed-photon-phonon. In our proposed photo-curing method, we use the dressed-photon-phonon for fixing heterogeneous QDs in a solution of a photo-curable polymer. In order to induce photo-curing in a self-assembled manner, the mixture in the solution, in which the QDs freely float due to thermal energy, is illuminated by assisting light.

An overview of our proposed method is schematically shown in Fig. 7.11. We assume a mixture containing two types of QDs, QD_A and QD_B, and a photo-curable polymer, and the mixture is irradiated with assisting light having photon energy $h\nu_{assist}$. The transition energies of QD_A and QD_B are $E_{A.bg}$ and $E_{B.bg}$, respectively, and the activation energy of the photo-curable polymer is $E_{poly.act}$. When the energies satisfy the condition $E_{A.bg} < h\nu_{assist} < E_{poly.act} < E_{B.bg}$, the following process can be induced. If the numbers of QDs, or in other words, their volume densities, in the mixture are not sufficiently high and they rarely encounter each other, only QD_A spontaneously emits visible light by absorbing the assisting light. In this case, no subsequent physical or chemical reaction occurs. On the other hand, if the density is sufficiently high that the QDs can frequently encounter each other, multistep energy excitation of the photo-curable polymer occurs due to dressed-photon-phonon interactions between neighboring QD_A and QD_B, and the photo-curable polymer is subsequently cured, as shown in the upper diagram in Fig. 7.11. Moreover, additional multistep excitation of QD_B by continuous illumination with the assisting light induces spontaneous emission of QD_B with a higher photon energy than the assisting light, and the photo-curable polymer in the surroundings of QD_A and QD_B

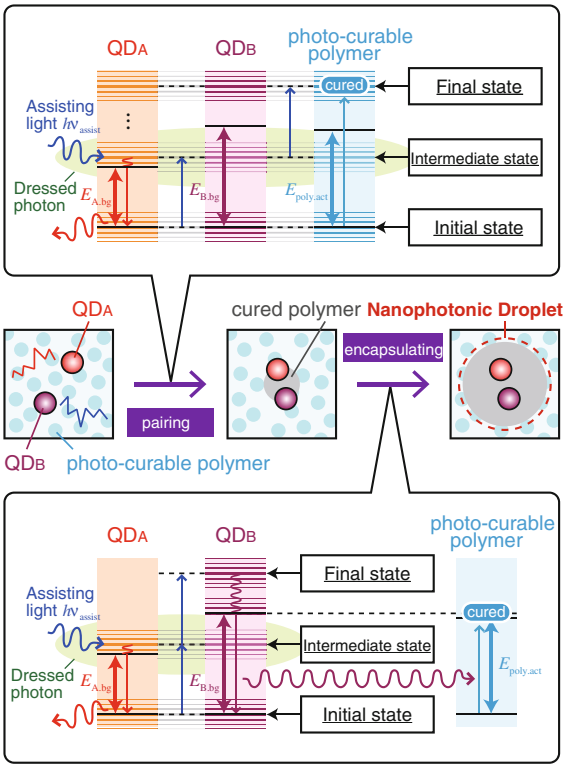


Fig. 7.11 Schematic diagram of fabrication process of a nanophotonic droplet based on phonon-assisted photo-curing method

is locally cured by absorption of this emitted light. As a result, the cured polymer gradually encapsulates the paired-QDs, preventing further combination or separation of the encapsulated QDs. Since the spatial distribution of the dressed-photon-phonon energy generated on the surface of the QDs is expressed by a Yukawa function [31], the separation between the two QDs is theoretically defined by the Yukawa function. Moreover, the encapsulated paired-QDs are necessarily composed of heterogeneous QDs, because the above sequence is induced only when QD_A and QD_B encounter each other. Descriptions of the electronic transitions via such a coupled state induced by the assisting light have been previously described in a recent report by the authors [39].

7.4.3 Experimental Demonstrations

First, we verified the effect of the phonon-assisted process by direct observation of a nanophotonic droplet. Figure 7.12 shows the energy conditions used for the demonstration. These conditions fulfill the previously described energy conditions

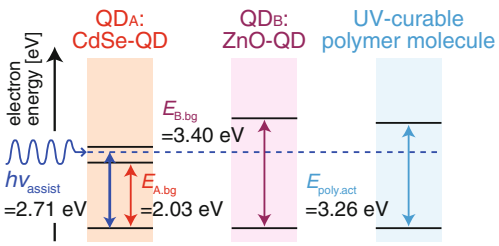


Fig. 7.12 Practical energy conditions of the materials used for the experimental verification: two heterogeneous QDs, CdSe-QDs and ZnO-QDs, and a UV-curable polymer

for inducing the sequential process of the phonon-assisted photo-curing. We used commercially available CdSe-QDs (Ocean Optics Inc.; *Evidot*) and ZnO-QDs prepared by sol-gel synthesis using photo-induced desorption [37]. The QD solutions were then dispersed in a UV-curable polymer (Norland Products Inc.; *NOA 65*) and illuminated by assisting light emitted from a 120 mW laser diode with a photon energy of 2.71 eV. The total amount of the mixed solution was limited to 50 μL to maintain spatially uniform illumination. This volume contained about 10^{14} CdSe-QDs and about 10^{12} ZnO-QDs. Under these experimental conditions, the QDs can be assumed to encounter each other at a sufficiently high frequency to induce the phonon-assisted photo-curing. After irradiation with the assisting light, the mixture was separated into cured and un-cured materials by centrifugation at 10,000 rpm for 5 min. The distilled cured material, which was assumed to contain nanophotonic droplets, was dispersed in a toluene solution and uniformly spin-coated on a Si substrate. Figure 7.13a, b show SEM images and microscope images of samples illuminated with assisting light for 30 and 90 min, respectively. With the shorter illumination time, as shown in Fig. 7.13a, a number of nanophotonic droplets of quite similar size were obtained. The diameter of each nanophotonic droplet was about 500 nm. The longer illumination time produced nanophotonic droplets of micrometer size, which were successfully observed in the microscope images, as shown in Fig. 7.13b. The images at the left- and right-hand sides in Fig. 7.13b show the sample under white light and UV light illumination, respectively. As shown in the right-hand image, each nanophotonic droplet clearly exhibited emission from the contained CdSe-QD.

The photoluminescence (PL) spectral properties of the produced nanophotonic droplets were investigated. By irradiating the sample with excitation light from a He-Cd laser, whose photon energy (3.82 eV) was sufficiently high to excite both the CdSe- and ZnO-QDs, the preferred optical energy transfer from ZnO-QDs to CdSe-QDs is expected to occur only in the case where these two are coupled. As a result of this transfer, the decreases and increases in the PL intensities from these QDs are expected to vary in an anti-correlated manner. That is, the PL intensity from the CdSe-QDs should increase while that from the ZnO-QDs decreases. We measured the relation between the irradiation time of the assisting light used for encapsulation and the PL intensities from both QDs. Because the surrounding polymer absorbs most of the PL light emitted from the excited state of an exciton in the ZnO-QDs, we measured the emission intensities from defect levels in the ZnO-QDs, which are assumed to be proportional to the emission from the excited state.

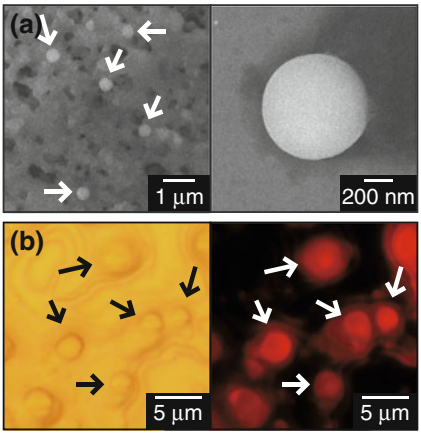


Fig. 7.13 **a** SEM images and **b** microscope images of mass-produced nanophotonic droplets

Figure 7.14b shows the obtained emission spectra of nanophotonic droplets with various illumination times of the assisting light. The left- and right-hand spectra, whose peak energies were lower and higher than 2.50 eV, represent PL from the lowest excitation level of the CdSe-QDs and the defect levels in the ZnO-QDs, respectively. The peak energies of the two spectra were 2.21 and 2.61 eV, respectively. As shown, the expected anti-correlated decreases and increases in the emission intensities from the CdSe- and ZnO-QDs were successfully observed. These results indicate that paired-QDs in each nanophotonic droplet selectively consisted of both CdSe- and ZnO-QDs, and that the number of paired-QDs and nanophotonic droplets gradually increased during irradiation with the assisting light.

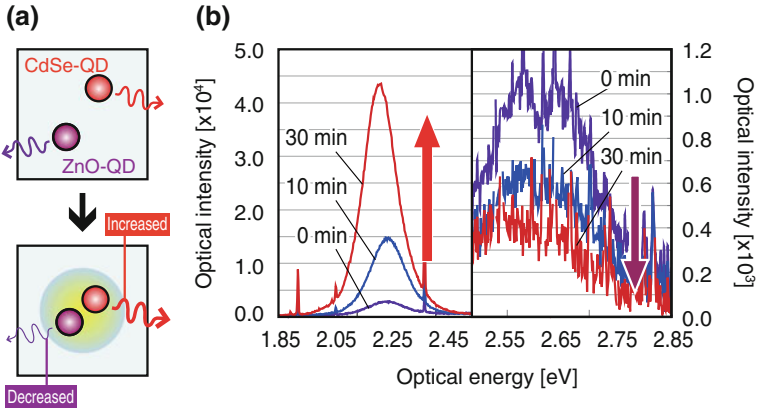


Fig. 7.14 **a** Schematic diagram of change of emission spectra due to pairing of CdSe- and ZnO-QDs. **b** Emission spectra of several samples formed with various illumination times of the assisting light

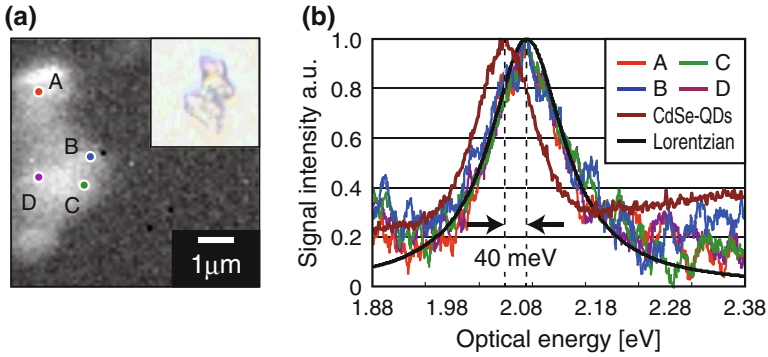


Fig. 7.15 **a** Obtained emission image of aggregated nanophotonic droplets. **b** Emission spectra of multiple nanophotonic droplets, collectively observed emission spectrum of CdSe-QDs, and a Lorentzian fitted curve

In order to compare multiple emission spectra of individual nanophotonic droplets and verify the uniformity of their optical properties, we constructed a two-dimensional map of the emission spectra of samples obtained with a micro-spectrophotometer (JASCO; NFGP-740) [41]. Samples were irradiated by a laser diode with a photon energy of 3.06 eV. The spatial resolution of the experimental setup was expected to be less than 200 nm. Figure 7.15a shows a two-dimensional emission image of aggregated nanophotonic droplets. The inset represents a bright-field image under white-light illumination. The individual emission spectra of nanophotonic droplets A, B, C, and D indicated in Fig. 7.15a are respectively plotted in Fig. 7.15b.

As shown, the heights and widths of the emission spectra are quite similar to each other. Because they can be simply fitted by a single Lorentzian, which is generally applied to approximate the emission spectrum of a single illuminant, it is considered that the nanophotonic droplets contain single CdSe-QDs of similar size and coupled ZnO-QDs. The results demonstrate the uniformity of the optical properties of mass-produced nanophotonic droplets. We also measured the collectively observed emission spectrum of multiple CdSe-QDs, as shown in Fig. 7.15b. The sample was prepared by mixing a solution of CdSe-QDs and a UV-curable polymer, which was directly spin-coated on a substrate without any irradiation with assisting light. There was a 40 meV difference between the peak optical energy in the spectra of the individual nanophotonic droplets and the collective spectrum of the multiple CdSe-QDs. Such a large difference is due to not only re-absorption of emission within the collective CdSe-QDs, but also size-selectivity due to the size-resonance between CdSe- and ZnO-QDs during the pairing process. The result indicates that ZnO-QDs preferably paired with smaller CdSe-QDs in the mixed solution and revealed such uniformity in their emission spectra.

7.4.4 Conclusion

We have described the concept of a nanophotonic droplet and its fabrication process based on a phonon-assisted self-organized photo-curing method consisting of pairing of heterogeneous QDs and encapsulation of paired-QDs. The proposed method utilizes a phonon-assisted excitation process between QDs and a photo-curable polymer. We verified the effectiveness of our technique via an experiment using sample solutions containing CdSe-QDs, ZnO-QDs, and a UV-curable polymer. Moreover, we observed the expected characteristic optical responses of the sample, demonstrating the validity of our idea.

We successfully verified the uniformity of the optical properties of mass-produced nanophotonic droplets. This is a fundamental feature in the concept of our nanometric optical system, because it will allow us to construct a number of functional nanometric setups with uniform optical properties by an all-autonomous fabrication process, which will enable direct application of our method to various fields where there is a demand for novel nanophotonics and nanotechnologies that will contribute to practical developments.

7.5 Summary

In this Chapter, we described the characteristics of the hierarchy of our nanometric optical system, which is implemented by exploiting some of the basic concepts of nanophotonics and dressed photon technology. All experimental demonstrations of nanometric optical system are designed to be conducted at the macro-scale. Moreover, such characteristic features of nanometric optical systems are realized not by using rare materials, but by utilizing novel design and engineering methods. These are important and fundamental issues from the standpoint of developing nanophotonics and dressed photon technology.

For further developments in nanometric optical systems, several related nanometric phenomena should continue to be looked at more deeply, such as the behavior of photons in nanometric space, excitation transfer and relaxation in nanometric space, and the primary mechanisms of phonon-assisted processes. Such fundamental studies are expected to contribute to the establishment of novel applications of nanometric optical systems in the near future.

Acknowledgments The author thanks the members of Ohtsu Research Group at the University of Tokyo, and Dr. M. Hoga, Mr. Y. Ohyagi, Ms. Y. Sekine, and Dr. H. Fujita in Dai Nippon Printing Co. Ltd. for their valuable contributions. Part of this work was supported by the Research and Development Program for Innovative Energy Efficiency Technology funded by the New Energy and Industrial Technology Development Organization (NEDO) of Japan.

References

1. H.H. Hoplins, Proc. Roy. Soc. **A217**, 408 (1953)
2. M. Ohtsu, Nanophotonics **1**, 83 (2012)
3. M. Naruse, T. Yatsui, W. Nomura, N. Hirose, M. Ohtsu, Opt. Exp. **13**, 9265 (2005)
4. M. Han, X. Gao, J.Z. Su, S. Nie, Nature Biotech. **19**, 631 (2001)
5. S.A. Crooker, J.A. Hollingsworth, S. Tretiak, V.I. Klimov, Phys. Rev. Lett. **89**, 186802 (2002)
6. M. Strassburg, M. Dworzak, H. Born, R. Heitz, A. Hoffmann, M. Bartels, K. Lischka, D. Schikora, J. Christen, Appl. Phys. Lett. **80**, 473 (2002)
7. T. Kawazoe, K. Kobayashi, M. Ohtsu, Appl. Phys. Lett. **86**, 103102 (2005)
8. T. Yatsui, S. Sangu, T. Kawazoe, M. Ohtsu, S.-J. An, J. Yoo, G.-C. Yi, Appl. Phys. Lett. **90**, 223110 (2007)
9. W. Nomura, T. Yatsui, T. Kawazoe, M. Ohtsu, J. Nanophotonics **1**, 011591 (2007)
10. T. Kawazoe, S. Tanaka, M. Ohtsu, J. Nanophotonics **2**, 029502 (2008)
11. M. Naruse, H. Hori, K. Kobayashi, T. Kawazoe, M. Ohtsu, Appl. Phys. B **102**, 717 (2011)
12. T. Kawazoe, M. Ohtsu, S. Aso, Y. Sawado, Y. Hosoda, K. Yoshizawa, K. Akahane, N. Yamamoto, M. Naruse, Appl. Phys. B **103**, 537 (2011)
13. H. Yonemitsu, T. Kawazoe, K. Kobayashi, M. Ohtsu, J. Photolumin. **122**, 230 (2007)
14. T. Yatsui, K. Hirata, W. Nomura, Y. Tabata, M. Ohtsu, Appl. Phys. B **93**, 55 (2008)
15. W. Nomura, T. Yatsui, Y. Yanase, K. Suzuki, M. Fujita, A. Kamata, M. Naruse, M. Ohtsu, Appl. Phys. B **99**, 75 (2010)
16. T. Kawazoe, H. Fujiwara, K. Kobayashi, M. Ohtsu, IEEE J. Sel. Top Quantum Electron. **15**, 1380 (2009)
17. H. Fujiwara, T. Kawazoe, M. Ohtsu, Appl. Phys. B **98**, 283 (2010)
18. M. Naruse, T. Yatsui, T. Kawazoe, Y. Akao, M. Ohtsu, IEEE Trans. Nanotech. **7**, 14 (2008)
19. N. Tate, W. Nomura, T. Yatsui, M. Naruse, M. Ohtsu, Opt. Exp. **16**, 607 (2008)
20. N. Tate, H. Tokoro, K. Takeda, W. Nomura, T. Yatsui, T. Kawazoe, M. Naruse, S. Ohkoshi, M. Ohtsu, Appl. Phys. B **98**, 685 (2010)
21. M. Naruse, T. Yatsui, T. Kawazoe, N. Tate, H. Sugiyama, M. Ohtsu, Appl. Phys. Exp. **1**, 112101 (2008)
22. N. Tate, H. Sugiyama, M. Naruse, W. Nomura, T. Yatsui, T. Kawazoe, M. Ohtsu, Opt. Exp. **17**, 11113 (2009)
23. N. Tate, M. Naruse, T. Yatsui, T. Kawazoe, M. Hoga, Y. Ohyagi, T. Fukuyama, M. Kitamura, M. Ohtsu, Opt. Exp. **18**, 7497 (2010)
24. N. Tate, M. Naruse, T. Yatsui, T. Kawazoe, M. Hoga, Y. Ohyagi, Y. Sekine, T. Fukuyama, M. Kitamura, M. Ohtsu, in Holograms—Recording Materials and Applications, ed. by I. Naydenova (InTech, Rijeka, 2011), p. 341
25. N. Tate, M. Naruse, W. Nomura, T. Kawazoe, T. Yatsui, M. Hoga, Y. Ohyagi, Y. Sekine, H. Fujita, M. Ohtsu, Opt. Exp. **19**, 18260 (2011)
26. C.R. Kagan, C.B. Murray, M.G. Bawendi, Phys. Rev. B Condens. Matter. **54**, 8633 (1996)
27. M. Achermann, M.A. Petruska, S.A. Crooker, V.I. Klimov, J. Phys. Chem. B **107**, 13782 (2003)
28. T. Franzl, D.S. Koktysh, T.A. Klar, A.L. Rogach, J. Feldmann, N. Gaponik, Appl. Phys. Lett. **84**, 2904 (2004)
29. T. Kawazoe, K. Kobayashi, J. Lim, Y. Narita, M. Ohtsu, Phys. Rev. Lett. **88**, 067404 (2002)
30. S. Sangu, K. Kobayashi, M. Ohtsu, J. Microsc. **202**, 279 (2001)
31. M. Ohtsu (ed.), *Progress in Nano-Electro-Optics II* (Springer, Berlin, 2004)
32. H.J. Calmichael, in *Statistical Methods in Quantum Optics 1* (Springer-Verlag, Berlin, 1999)
33. M. Naruse, H. Hori, K. Kobayashi, P. Holmström, L. Thylén, M. Ohtsu, Opt. Exp. **18**, A544 (2010)
34. T. Kawazoe, K. Kobayashi, S. Sangu, M. Ohtsu, Appl. Phys. Lett. **82**, 2957 (2003)
35. M. Naruse, K. Leibnitz, F. Peper, N. Tate, W. Nomura, T. Kawazoe, M. Murata, M. Ohtsu, Nano. Commun. Netw. **2**, 189 (2011)
36. S. Yukutake, T. Kawazoe, T. Yatsui, W. Nomura, K. Kitamura, M. Ohtsu, Appl. Phys. B **99**, 415 (2010)

37. Y. Liu, T. Morishima, T. Yatsui, T. Kawazoe, M. Ohtsu, Nanotechnology **22**, 215605 (2011)
38. T. Kawazoe, M.A. Mueed, M. Ohtsu, Appl. Phys. B **104**, 747 (2011)
39. N. Tate, Y. Liu, T. Kawazoe, M. Naruse, T. Yatsui, M. Ohtsu, Appl. Phys. B (2012). doi:10.1007/s00340-012-5249-5
40. K. Kitamura, T. Kawazoe, M. Ohtsu, Appl. Phys. B **107**, 293 (2012)
41. N. Tate, Y. Liu, T. Kawazoe, M. Naruse, T. Yatsui, M. Ohtsu, Appl. Phys. B (2012). doi:10.1007/s00340-012-5285-1

Chapter 8
Photonic Neuromorphic Signal Processing and Computing

Alexander N. Tait, Mitchell A. Nahmias, Yue Tian, Bhavin J. Shastri and Paul R. Prucnal

Abstract There has been a recent explosion of interest in spiking neural networks (SNNs), which code information as spikes or events in time. Spike encoding is widely accepted as the information medium underlying the brain, but it has also inspired a new generation of neuromorphic hardware. Although electronics can match biological time scales and exceed them, they eventually reach a bandwidth fan-in trade-off. An alternative platform is photonics, which could process highly interactive information at speeds that electronics could never reach. Correspondingly, processing techniques inspired by biology could compensate for many of the shortcomings that bar digital photonic computing from feasibility, including high defect rates and signal control problems. We summarize properties of photonic spike processing and initial experiments with discrete components. A technique for mapping this paradigm to scalable, integrated laser devices is explored and simulated in small networks. This approach promises to wed the advantageous aspects of both photonic physics and unconventional computing systems. Further development could allow for fully scalable photonic networks that would open up a new domain of ultrafast, robust, and adaptive processing. Applications of this technology ranging from nanosecond response control systems to fast cognitive radio could potentially revitalize specialized photonic computing.

8.1 Introduction

The brain, unlike the von Neumann processors found in conventional computers, is very power efficient, extremely effective at certain computing tasks, and highly adaptable to novel situations and environments. While these favorable properties are often largely credited to the unique connectionism of biological nervous systems, it is

A. N. Tait · M. A. Nahmias · Y. Tian · B. J. Shastri · P. R. Prucnal (✉)
Lightwave Communications Laboratory, Department of Electrical Engineering,
Princeton University, Princeton, NJ 08544, USA
e-mail: prucnal@princeton.edu

M. Naruse (ed.), *Nanophotonic Information Physics*,
Nano-Optics and Nanophotonics, DOI: 10.1007/978-3-642-40224-1_8,

clear that the cellular dynamics of individual neurons also play an indispensable role in attaining these performance advantages. On the cellular level, neurons operate on information encoded as events or *spikes*, a type of signal with both analog and digital properties. Spike processing exploits the efficiency of analog signals while overcoming the problem of noise accumulation inherent in analog computation, which can be seen as a first step to attaining the astounding capabilities of bio-inspired processing [1]. Other physical and signal processing features at the single neuron level, including hybrid analog-digital signals, representational interleaving, co-location of memory and processing, unsupervised statistical learning, and distributed representations of information, have been implicated in various positive and conventionally impossible signal processing capabilities. Engineers have tried for nearly half a century to take inspiration from neuroscience by separating key properties that yield performance from simple idiosyncrasies of biology.

If some biological properties are distilled while others deemed irrelevant, then a bio-inspired engineered system could also incorporate nonbiological properties that may lead to computational domains that are potentially unexplored and/or practically useful. Many microelectronic platforms have attempted to emulate some advantages of neuron-like architectures while incorporating techniques in technology; however, the majority target rather than exceed biological time scales. What kind of signal processing would be possible with a bio-inspired visual front-end that operates ten million times faster than its biological counterpart? Unfortunately, microelectronic neural networks that are both fast and highly interconnected are subject to a fundamental bandwidth connection-density tradeoff, keeping their speed limited.

Photonic platforms offer an alternative approach to microelectronics. The high speeds, high bandwidth, and low cross-talk achievable in photonics are very well suited for an ultrafast spike-based information scheme with high interconnection densities. In addition, the high wall-plug efficiencies of photonic devices may allow such implementations to match or eclipse equivalent electronic systems in low energy usage. Because of these advantages, photonic spike processors could access a *picosecond, low-power* computational domain that is inaccessible by other technologies. Our work aims to synergistically integrate the underlying physics of photonics with bio-inspired spike-based processing. This novel processing domain of *ultrafast cognitive computing* could have numerous applications where quick, temporally precise and robust systems are necessary, including: adaptive control, learning, perception, motion control, sensory processing, autonomous robotics, and cognitive processing of the radio frequency spectrum.

In this chapter, we consider the system implications of wedding a neuron-inspired computational primitive with the unique device physics of photonic hardware. The result is a system that could emulate neuromorphic algorithms at rates millions of times faster than biology, while also overcoming both the scaling problems of digital optical computation and the noise accumulation problems of analog optical computation by taking inspiration from neuroscience. Our approach combines the picosecond processing and switching capabilities of both linear and nonlinear optical device technologies to integrate both analog and digital optical processing into a single hardware architecture capable of ultrafast computation without the need for

conversion from optical signals to electrical currents. With this hybrid analog-digital processing primitive, it will be possible to implement processing algorithms of far greater complexity than is possible with existing optical technologies, and far higher bandwidth than is possible with existing electronic technologies.

The design and implementation of our devices is based upon a well-studied and paradigmatic example of a hybrid computational primitive: the *spiking neuron*. These simple computational elements integrate a small set of basic operations (delay, weighting, spatial summation, temporal integration, and thresholding) into a single device which is capable of performing a variety of computations depending on how its parameters are configured (e.g., delays, weights, integration time constant, threshold). The *leaky-integrate-and-fire* (LIF) neuron model is a mathematical model of the spiking dynamics which pervade animal nervous systems, well-established as the most widely used model of biological neurons in theoretical neuroscience for studying complex computation in nervous systems [2]. LIF neurons have recently attracted the attention of engineers and computer scientists for the following reasons:

- Algorithmic Expressiveness: They provide a powerful and efficient computational primitive from the standpoint of computational complexity and computability theory;
- Hardware Efficiency/Robustness: They are both robust and efficient processing elements. Complex algorithms can be implemented with less hardware;
- Utility in signal processing: There are known pulse processing algorithms for complex signal processing tasks (drawn from neuro-ethology). Pulse processing is already being used for implementing robust sensory processing algorithms in analog VLSI.

After comparing the fundamental physical traits of electronic and photonic platforms for implementation of neuromorphic architectures, this chapter will present a detailed cellular processing model (i.e. LIF) and describe the first emulation of this model in ultrafast fiber-optic hardware. Early experiments in photonic learning devices, critical for controlling the adaptation of neural networks, and prototypical demonstrations of simple bio-inspired circuits will be discussed. Recent research has sought to develop a scalable neuromorphic primitive based on the dynamics of semiconductor lasers which could eventually be integrated in very large lightwave neuromorphic systems on a photonic chip. This chapter will focus on the computational primitive, how it can be implemented in photonics, and how these devices interact with input and output signals; a detailed account of the structure and implementation of a network that can link many of these devices together is beyond the scope of this chapter.

8.2 Neuromorphic Processing in Electronics and Photonics

The harmonious marriage of a processing paradigm to the physical behaviors that bear it represents an important step in efficiency and performance over approaches that aim to abstract physics away entirely. Both thalamocortical structures observed

in biology and emerging mathematical models of network information integration exhibit a natural separation of computation and communication. Computation rarely takes place in the axons of a neuron, but these communication pathways and their rich configurability are crucial for the informatic complexity of the network as a whole. Because of their favorable communication properties (i.e. low dispersion, high bandwidth, and low cross-talk), photonic spiking network processors could access a computationally rich and high bandwidth domain that is inaccessible by other technologies. This domain, which we call *ultrafast cognitive computing*, represents an unexplored processing paradigm that could have a wide range of applications in adaptive control, learning, perception, motion control, sensory processing (vision systems, auditory processors, and the olfactory system), autonomous robotics, and cognitive processing of the radio frequency spectrum.

Currently developing cortically-inspired microelectronic architectures including IBM's neurosynaptic core [3, 4] and HP's proposed memristive nanodevices [5, 6] use a dense mesh of wires called a crossbar array to achieve heightened network configurability and fan-in, which are less critical in conventional architectures. These architectures aim to target clock rates comparable to biological time scales rather than exceed them. At high-bandwidths, however, densely packed electronic wires cease to be effective for communication. Power use increases drastically, and signals quickly attenuate, disperse, or couple together unfavorably, especially on a crossbar array, which has a large area of closely packed signal wires. In contrast, photonic channels can support the high bandwidth components of spikes without an analogous speed, power, fan-in, and cross-talk trade-off.

Optical neural networks based on free space and fiber components have been explored for interconnection in the past, but undeveloped fiber-optic and photonic technologies in addition to scalability limitations of particular strategies (e.g. diffraction limit, analog noise accumulation) relegated these systems to small laboratory demonstrations [7–9]. Optical processing elements and lasers were not realistically available for use as fast computational primitives, so the ultrafast domain where optics fundamentally dominates microelectronics was never fully considered. Perhaps as culpable in the initial failure of optical neural networks was the sparsity of neurocomputational theories based on spike timing. The current states of photonic technology and computational neuroscience have matured enormously and are ripe for a new breed of photonic neuromorphic research.

8.3 Photonic Spike Processing

Optical communication has extensively utilized the high bandwidth of photonics, but, in general, approaches to optical computing have been hindered by scalability problems. We hypothesize that the primary barrier to exploiting the high bandwidth of photonic devices for computing lies in the model of computation being used, not solely in the performance, integration, or fabrication of the devices. Many years of research have been devoted to photonic implementation of traditional models of

computation, yet neither analog nor digital approaches has proven scalable due primarily to challenges of cascadability and fabrication reliability. Analog photonic processing has found widespread application in high bandwidth filtering of microwave signals, but the accumulation of phase noise, in addition to amplitude noise, makes cascaded operations particularly difficult. Digital logic gates that suppress noise accumulation have also been realized in photonics, but photonic devices have not yet met the extremely high fabrication yield required for complex digital operations, a tolerance requirement that is greatly relaxed in systems capable of biomorphic adaptation. In addition, schemes that take advantage of the multiple available wavelengths require ubiquitous wavelength conversion, which can be costly, noisy, inefficient, and complicated.

The optical channel is highly expressive and correspondingly very sensitive to phase and frequency noise. Any proposal for a computational primitive must address the issue of practical cascadability, especially if multiple wavelength channels are intended to be used. Our proposed unconventional computing primitive addresses the traditional problem of noise accumulation by interleaving physical representations of information. Representational interleaving, in which a signal is repeatedly transformed between coding schemes (digital-analog) or physical variables (electronic-optical), can grant many advantages to computation and noise properties. For example, a logarithm transform can reduce a multiplication operation to a simpler addition operation. As discussed in Sect. 8.3.1, the spiking model found in biology naturally interleaves robust, discrete representations for communication signals with precise, continuous representations for computation variables in order to leverage the benefits of both types of coding. It is natural to deepen this distinction to include physical representational aspects, with the important result that optical noise does not accumulate. When a pulse is generated, it is transmitted and routed through a linear optical network with the help of its wavelength identifier. It is received only once into a fermionic degree of freedom, such as a deflection of carrier concentration in a photonic semiconductor or a current pulse in a photodiode. The soma tasks occur in a domain that is in some ways computationally richer (analog) and in other ways more physically robust (incoherent) to the type of phase and frequency noise that can doom optical computing architectures.

Another major hurdle faced by substantial photonic systems is relatively poor device yield. While difficult for digital circuits to build in redundancy overhead, neuromorphic systems are naturally reconfigurable and adaptive. Biological nervous systems adapt to shunt signals around faulty neurons without the need for external control. The structural fluidity of neural algorithms will itself provide an effective resistance against fabrication defects.

8.3.1 Spiking Signals

On the cellular level, the brain encodes information as events or *spikes* in time [10], "hybrid signals" with both analog and digital properties as illustrated in Fig. 8.1.

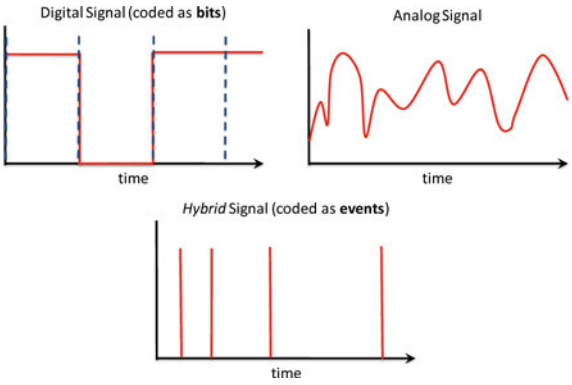


Fig. 8.1 Spiking neural networks encode information as events in time rather than bits. Because the time at which a spike occurs is analog while its amplitude is digital, the signals use a mixed-signal or hybrid encoding scheme

	Digital	Analog	Spiking
Robustness			
Expressiveness			
Power Efficiency			
Bandwidth limiter	Clock speed	Minimum SNR	Pulse width

Fig. 8.2 Comparison of digital, analog, and hybrid systems in terms of robustness, expressiveness, power efficiency, and bandwidth limiter

Spike processing has evolved in biological (nervous systems) and engineered (neuromorphic analog VLSI) systems using LIF neurons as processing primitives that encode data in the analog timing of spikes as inputs. Through its fine grained interleaving of analog and digital processing, it provides the means for achieving scalable high bandwidth photonic processing by overcoming both the analog noise accumulation problem [1] and the bandwidth inefficiency of digital processing. These unique properties of the spike processing model of computation enable a hybrid analog and digital approach that will allow photonics hardware to scale its processing complexity and efficiency. The foregoing discussion is summarized in Fig. 8.2 which compares digital, analog, and hybrid systems in terms of robustness, expressiveness, power efficiency, and the characteristic property that determines usable bandwidth. Spiking signals, which in addition to being inherently advantageous, carry information in a natural and accessible fashion that forms the very foundation of some of the astounding capabilities of systems studied in neuroscience.

8.3.2 Spike Processor: Computational Primitive

Like a gate in a digital computer, a spike processor serves as the baseline unit of larger interconnected networks that can perform more complex computations. There are five key computational properties to consider: (1) *integration*, the ability to sum and integrate both positive and negative weighted inputs over time; (2) *thresholding*,

the ability to make a decision whether or not to send a spike; (3) *reset*, the ability to have a small refractory period during which no firing can occur immediately after a spike is released; (4) *pulse generation*, the ability to generate new pulses; and (5) *adaptability*, the ability to modify and regulate response properties on slow timescales based on statistical properties of environmental inputs and/or training.

These five properties all play important roles in the emulation of the widely accepted neurocomputational primitive, the LIF neuron. This model is more computationally powerful than either the rate or the earlier perceptron models [11], and can serve as the baseline unit for many modern cortical algorithms [12–14]. Although not every property is needed to do constrained useful computations, each one serves an important purpose to assure a high level of richness and robustness in the overall computational repertoire.

Integration is a temporal generalization of summation in older perceptron-based models. Spikes with varying amplitudes and delays arrive at the integrator, changing its state by an amount proportional to their input amplitudes. Excitatory inputs increase its state, while inhibitory inputs deplete it. The integrator is especially sensitive to spikes that are closely clustered in time or with high amplitudes, the basic functionality of temporal summation. Eventually, the state variable will decay to its equilibrium value without any inputs. Both the *amplitude* and *timing* play an important role for integration.

Thresholding determines whether the output of the integrator is above or below a predetermined threshold value, T. The neuron makes a decision based on the state of the integrator, firing if the integration state is above the threshold. Thresholding is the center of nonlinear decision making in a spiking system, which reduces the dimensionality of incoming information in a useful fashion and plays a critical role in cleaning up amplitude noise that would otherwise cause the breakdown of analog computation.

The reset condition resets the state of the integrator to a low, rest value immediately after a spike processor fires, causing a *refractory period* in which it is impossible or difficult to cause the neuron to fire again. It plays the same role in time as the thresholder does for amplitude, cleaning up *timing jitter* and preventing the temporal spreading of excitatory activity while putting a bandwidth cap on the output of a given spiking unit. It is additionally a necessary component for simulating the rate model of neural computation with spiking neurons.

Pulse generation refers to the ability for a system to spontaneously generate pulses. If pulses are not regenerated as they travel through systems, they will eventually be lost in noise. A system with this property can generate pulses without the need to trigger on input pulses whenever the integrator's state variable reaches the threshold, T, which is crucial for simulating the rate model.

Adaptability is the network's ability to change to better suit changing environmental and system conditions. Adaptation of network parameters typically occurs on time scales much slower than spiking dynamics and can be separated into either unsupervised learning, where alterations in overall signal statistics cause automatic adjustments, or supervised learning, where changes are guided based on the behavior of the system compared to a desired behavior presented by a teacher.

Table 8.1 Spiking neural automata properties

Integration	Temporally sum weighted inputs
Thresholding	Fire one spike when integrator state exceeds some level
Reset	Return integrator state to rest immediately following a spike
Pulse generation	Introduce a new spike into the network
Adaptability	Modify behavioral parameters based on input statistics

Because of the extreme reconfigurability and fluidity of massively parallel neural networks, adaptation rules are necessary to stabilize the system's structure and accomplish a desired task. Adaptation also corrects catastrophic system alterations by, for example, routing signals around a failed neuron to maintain overall process integrity.

The characteristics mentioned above are summarized in Table 8.1. In summary, (1) and (4) provide key properties of a system that is the basis of asynchronous communication, while (2) and (3) clean up amplitude and temporal noise, respectively, to allow for cascadability. Any processor designed to closely emulate LIF neurons should have all five properties.

In photonics, the challenge of creating a flexible, scalable, and efficient native hardware implementation of the spike processing model of computation lies in the design of the critical, programmable component devices of the LIF neuron. The ability of the LIF to perform matched filtering, dimensionality reduction, evidence accumulation, decision making, and communication of results requires analog, tunable component hardware elements as described in Table 8.2. These parts must perform adjustable operations and demonstrate configurable interconnectivity between large numbers of LIF neurons.

The weights and delays of the matched filter are applied at the interconnection of two LIF neurons and form a complicated junction that acts as the primary communications between the spiking elements. These interconnection controls are a key feature of the LIF and must be capable of rapid read and write reconfigurability and adjustability either through direct programming or through learning algorithms like spike timing dependent plasticity (STDP).

In the next section we review the LIF neuron model which is a mathematical model of the spiking neuron primitive.

Table 8.2 Fabricating LIF neurons requires photonic elements

Challenges	Required analog control	Photonic device candidates
Matched filtering	Connectivity weights and delays	Amplifiers, tunable microring resonators
Dimensionality reduction	Summation of input lines	Couplers, waveguides
Evidence accumulation	Temporal integration constant	Photodiodes
Decision making	Adjustable thresholding	Laser bias; saturable absorber devices

8.4 Spiking Neuron Model

Our devices are based upon a well-studied and paradigmatic example of a hybrid computational primitive: the spiking neuron. Studies of morphology and physiology have pinpointed the LIF model as an effective spiking model to describe a variety of different biologically observed phenomena [2]. From the standpoint of computability and complexity theory, LIF neurons are powerful and efficient computational primitives that are capable of simulating both Turing machines and traditional sigmoidal neural networks [15]. These units perform a small set of basic operations (delaying, weighting, spatial summation, temporal integration, and thresholding) that are integrated into a single device capable of implementing a variety of processing tasks, including binary classification, adaptive feedback, and temporal logic.

The basic biological structure of a LIF neuron is depicted in Fig. 8.3a. It consists of a dendritic tree that collects and sums inputs from other neurons, a soma that acts as a low pass filter and integrates the signals over time, and an axon that carries an action potential, or spike, when the integrated signal exceeds a threshold. Neurons are connected to each other via synapses, or extracellular gaps, across which chemical signals are transmitted. The axon, dendrite, and synapse all play an important role in the weighting and delaying of spike signals.

According to the standard LIF model, neurons are treated as an equivalent electrical circuit. The membrane potential $V_m(t)$, the voltage difference across their membrane, acts as the primary internal (activation) state variable. Ions that flow across the membrane experience a resistance $R = R_m$ and capacitance $C = C_m$ associated

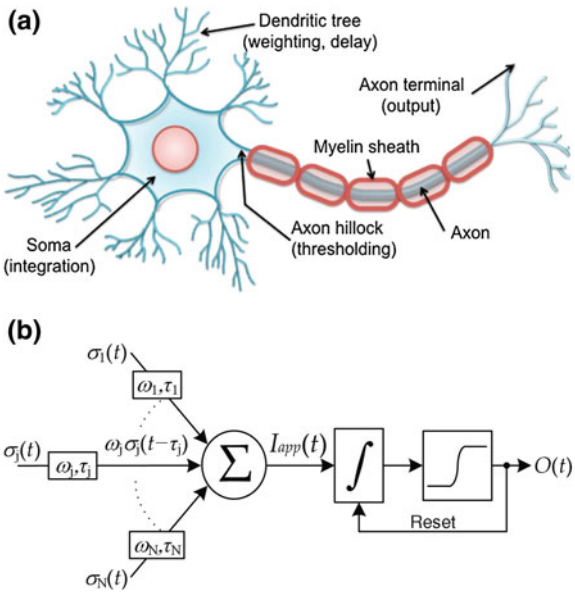


Fig. 8.3 **a** Schematic and **b** operation of a leaky integrate-and-fire neuron. Weighted and delayed input signals are summed onto the soma, which integrates them together and makes a spike or no-spike decision. The resulting spike is sent to other neurons in the network

with the membrane. The soma is effectively a first-order low-pass filter, or a leaky integrator, with the integration time constant $\tau_m = R_m C_m$ that determines the exponential decay rate of the impulse response function. The leakage current through R_m drives the membrane voltage $V_m(t)$ to 0, but an active membrane pumping current counteracts it and maintains a resting membrane voltage at a value of $V_m(t) = V_L$.

Figure 8.3b shows the standard LIF neuron model [15]. A neuron has: (1) N inputs which represent induced currents through input synapses $\sigma_j(t)$, that are continuous time series consisting either of spikes or continuous analog values; (2) an internal activation state $V_m(t)$; and (3) a single output state $O(t)$. Each input is independently weighted[1] by ω_j and delayed by τ_j resulting in a time series that is spatially summed (summed pointwise). This aggregate input electrical current, $I_{\text{app}}(t) = \Sigma_{j=1}^{n} \omega_j \sigma_j(t - \tau_j)$. The result is then temporally integrated using an exponentially decaying impulse response function resulting in the activation state $V_m(t) = V_L e^{\frac{t-t_0}{\tau_m}} + \frac{1}{C_m} \int_0^{t-t_0} I_{\text{app}}(t-s) e^{\frac{s}{\tau_m}} ds$, where t_0 is the last time the neuron spiked. If $V_m(t) \geq V_{\text{thresh}}$, then the neuron outputs a spike, $O(t) = 1$, and $V_m(t)$ is set to V_{reset}. After issuing a spike, there is a short period of time, the refractory period, during which it is difficult to issue another spike; that is, if $O(t) = 1$ then $O(t - \Delta t) = 0$, $\Delta t \leq T_{\text{refract}}$. Consequently, the output of the neuron consists of a continuous time series comprised of spikes.

The parameters determining the behavior of the device are: the weights ω_j, delays τ_j, threshold V_{thresh}, resting potential V_L, refractory period T_{refract}, and the integration time constant τ_m. There are three influences on $V_m(t)$: passive leakage of current, an active pumping current, and external inputs generating time-varying membrane conductance changes. These three influences are the three terms contributing to the differential equation describing $V_m(t)$ as

$$\underbrace{\frac{dV_m(t)}{dt}}_{\text{Activation}} = \underbrace{\frac{V_L}{\tau_m}}_{\text{Active pumping}} - \underbrace{\frac{V_m(t)}{\tau_m}}_{\text{Leakage}} + \underbrace{\frac{1}{C_m} I_{app}(t)}_{\text{External input}}; \tag{8.1}$$

$$\text{if } V_m(t) > V_{\text{thresh}} \text{ then release a pulse and set } V_m(t) \rightarrow V_{\text{reset}}. \tag{8.2}$$

8.5 Photonic Neuron Bench-Top Model

The first emulation of spiking LIF behavior in ultrafast optical components was a large fiber-based system that took up the area of an optical bench. Although this rudimentary photonic neuron was bulky, power hungry (2 W), and inefficient (1 %), it demonstrated a variety of important hybrid spike processing functions, including integration and thresholding [16–18]. Many of the same physical principles explored in these initial prototypes remain fundamental to more recently developed integrated models (Sect. 8.8).

[1] Since ω_j can either be positive or negative, both, excitation and inhibition can be implemented.

The most striking resemblance between the physics of photonics and the neural computing paradigm is shown in a direct correspondence between the equations governing the intracellular potential of a LIF neuron (see 8.1a) and the gain dynamics of a semiconductor optical amplifier (SOA):

$$\underbrace{\frac{dN'(t)}{dt}}_{\text{Activation}} = \underbrace{\frac{N'_{rest}}{\tau_e}}_{\text{Active pumping}} - \underbrace{\frac{N'(t)}{\tau_e}}_{\text{Leakage}} + \underbrace{\frac{\Gamma a(\lambda)}{E_p} N'(t) P(t)}_{\text{External input}} \tag{8.3}$$

where $N'(t)$ is the excited carrier concentration in the optically-coupled semiconductor junction, τ_e is the carrier lifetime, $P(t)$ is the incident optical power, and $\frac{\Gamma a(\lambda)}{E_p}$ is a wavelength-dependent light-matter coupling coefficient. The correspondence between the equations indicates that the natural physical behaviors of this standard electro-optical material can be made to emulate the primary analog portion of neural dynamics: summing and integration. Despite being dynamically isomorphic, these two equations operate on vastly different time scales; whereas the time constant $R_m C_m = \tau_m$ in biological neurons is around 10 ms, τ_e in SOAs typically falls within the range of 25–500 ps. It is thus possible to emulate a model of the neural integration function and perhaps tap into neuro-inspired processing techniques for applications at time scales millions to billions of times faster.

Electronic free carriers in the active region of the SOA are analogous to neuron state. Input spikes rapidly perturb the carrier concentration, which then decays back to its resting value. Optical inputs can perturb the SOA state either positively or negatively depending on their wavelength [19], which is represented in (8.3) by the gain parameter $a(\lambda)$. Wavelengths within the gain band of the SOA will deplete the carrier concentration while shorter wavelengths can pump it for the opposite effect. Figure 8.4 shows an experimental demonstration of simultaneous optical excitation and inhibition in an SOA integrator.

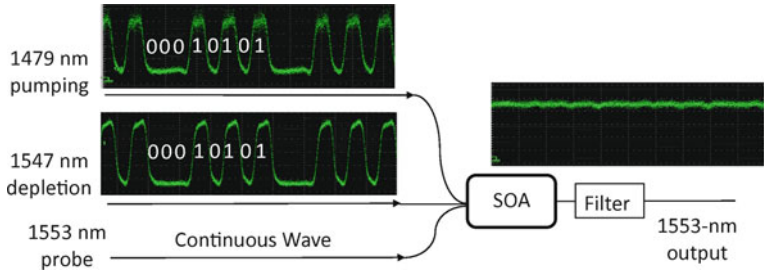


Fig. 8.4 Experimental setup for demonstrating simultaneous excitatory and inhibitory stimuli in the SOA through gain pumping and depletion by different wavelengths. Inputs have identical bit patterns, resulting in no net change in the probe power. Unbalanced inputs would result in positive and negative modulations of the SOA transmittance at probe wavelength

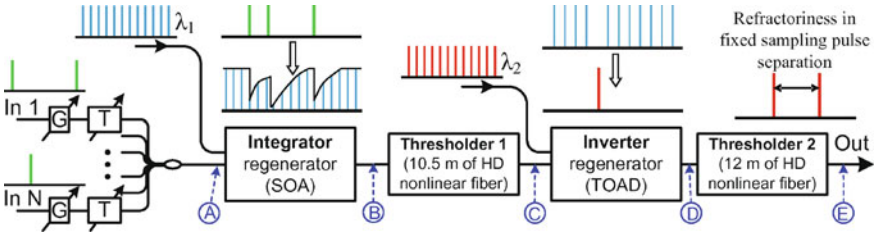


Fig. 8.5 Block diagram of a bench-top fiber based photonic neuron. *G* variable attenuator. *T* tunable delay line. λ_1 low power pulse train samples the carrier concentration of the SOA, which is depleted by input pulses

To extract information about the carrier concentration of the integrating SOA, a probe signal (λ_1 in Fig. 8.5) is modulated by its instant gain, which depends on the free carrier concentration. The stream of modulated pulses is sent to an optical thresholder, which makes the nonlinear spike or no-spike decision seen in biological spiking neurons. This represents the digital portion of the hybrid nature of the computation. Thresholding is performed by a nonlinear optical loop mirror (NOLM), which utilizes the nonlinear Kerr effect to induce an approximately sigmoid-like power transmission function [20]. Two NOLMs in succession were used in the experimental set-up shown in Fig. 8.5 to improve the sharpness of the thresholding characteristic.

Since both the SOA and the mode-locked laser are integratable devices, the ability to integrate the feedforward neuron depends crucially on shrinking the thresholder into a small footprint. A device has been recently been invented to better performs this task: the dual resonator enhanced asymmetric Mach-Zehnder interferometer (DREAM) [21, 22]. By tailoring the finesse of rings in each arm, the pulse width and peak power can be balanced to obtain an energy transfer response that very closely resembles the desired step function (Fig. 8.6). The DREAM could exceeed NOLM performance by four orders of magnitude in each of its key performance criteria:

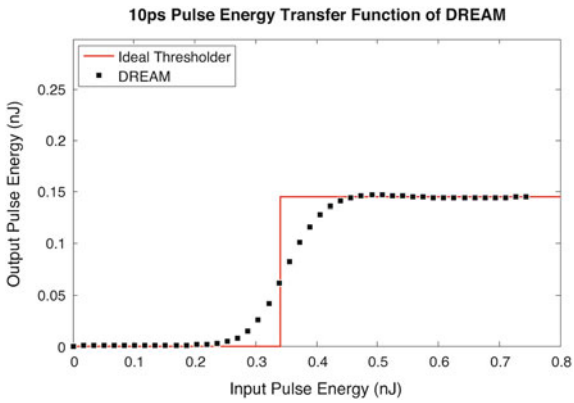


Fig. 8.6 Thresholding capabilities of DREAM simulated in a commercial FDTD environment. Input pulse widths are 10 ps, threshold level is 350 pJ. The ideal thresholding characteristic is a Heaviside step function (*red line*)

size, decision latency, and switching power. Decision latencies on the order of ten picoseconds are attainable using current material technologies [23].

Although this model successfully performs two qualitative features of biological computation integration and thresholding, it lacks a reset condition, the ability to generate optical pulses, and truly asynchronous behavior. Several modifications to this bench-top model, including a delayed output pulse fed back to reset the SOA, make it suitable for preliminary experiments in learning (Sect. 8.7) and simple lightwave neuromorphic circuits (Sect. 8.6). The original fiber-based photonic neuron is hardly scalable to networks of many neurons, but it identifies a new domain of ultrafast cognitive computing, which has informed the development of more advanced photonic neuron devices based on excitable laser dynamics, a model that is described in Sect. 8.8.

8.6 Lightwave Neuromorphic Circuits

Neuromorphic engineering provides a wide range of practical computing and signal processing tools by exploiting the biophysics of neuronal computation algorithms. Existing technologies include analog very-large-scale integration front-end sensor circuits that replicate the capabilities of the retina and the cochlea [24]. To meet the requirements of real-time signal processing, lightwave neuromorphic signal processing can be utilized to provide the high-speed and low-latency performance that is characteristic of photonic technology. We have demonstrated several small-scale lightwave neuromorphic circuits to mimic important neuronal behavior based on the bench-top model of the photonic neuron detailed in Sect. 8.5. Here, we present several prototypical lightwave neuromorphic circuits including: (1) simple auditory localization inspired by the barn owl [25], useful for light detection and ranging (LIDAR) localization; (2) the crayfish tail-flip escape response [18], which demonstrates accurate, picosecond pattern classification; and (3) principle and independent component analysis, which can adaptively separate mixed and/or corrupted signals based on statistical relationships between multiple sensors.

8.6.1 Barn Owl Auditory Localization Algorithm

Figure 8.7 shows a simple diagram of auditory localization. Due to the difference in position of object 1 and object 2, there is a time difference between the signals arriving at the owl's left sensor and right sensor, denoted as $\Delta T_1 = (t_{1a} - t_{1b})$ for object 1 and $\Delta T_2 = (t_{2a} - t_{2b})$ for object 2. Thus, the neuron can be configured to respond to a certain object location by adjusting the weight and delay of the neuron inputs. If the weighted and delayed signals are strong enough and arrive within the integration window, a given neuron spikes; otherwise no spike results and a different neuron corresponding to another location may fire. Figure 8.7 illustrates the

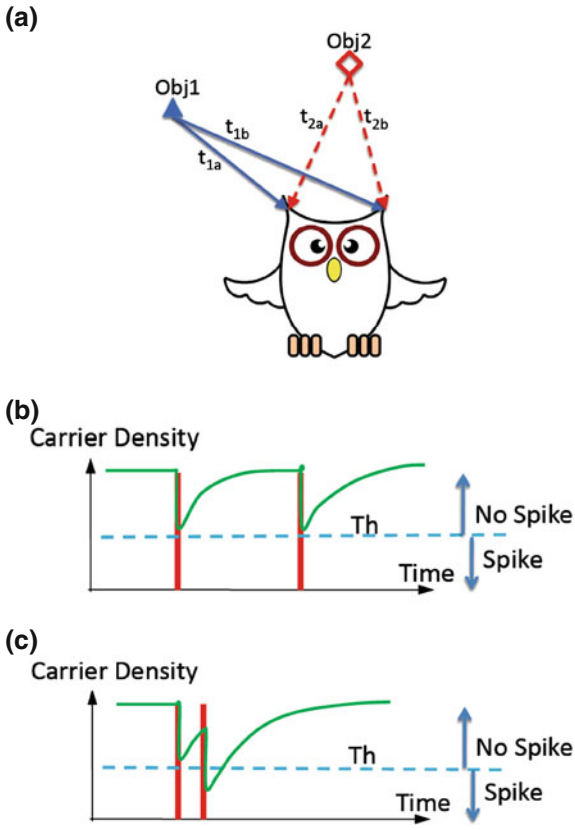


Fig. 8.7 **a** Schematic illustration of the auditory localization algorithm of the barn owl. **b** and **c** SOA carrier density when two signals are far apart (*no spike*) and when two signals are close (*spike*), respectively

corresponding SOA-based integrator response when the two weighted and delayed signals are relatively far apart. The stimulated signal cannot pass through the thresholder and therefore no spike is obtained. When the two inputs are close enough, the carrier density reaches the threshold and leads to a spike as depicted in Fig. 8.7. This algorithm could be used for front-end spatial filtering of RADAR waveform or LIDAR signals.

Figure 8.8 depicts the temporal sensitivity of the spike processor [17]. More specifically, Fig. 8.8 corresponds to the inputs consisting of a number of pulses (signals) with the same intensity but with different time interval, that is, in case I the input signals are close together (measured temporal resolution limited by the bandwidth of the photodetector) and in case II the input signals are further apart. After temporal integration at the SOA and thresholding at the optical thresholder, a spiking output is obtained as shown in Fig. 8.8. Due to the gain depletion property of the SOA, the spike output is inverted. No spike is observed when the input signals are close together (case I), while spike are observed when the input signals are further apart (case II).

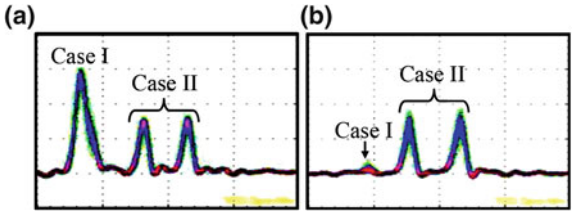


Fig. 8.8 **a** Input to the photonic neuron: (1) *case I* input signals are close together; and (2) *case II* input signals are further apart. **b** Photonic neuron outputs (inverted): (1) *case II* no spike when signals are close; and (2) *case II* spike when signals are further apart

8.6.2 Crayfish Tail-Flip Escape Response

We have also demonstrated a device for signal feature recognition based on the escape response neuron model of a crayfish [26]. Crayfish escape from danger by means of a rapid escape response behavior. The corresponding neural circuit is configured to respond to appropriately sudden stimuli. Since this corresponds to a life-or-death decision for the crayfish, it must be executed quickly and accurately. A potential application of the escape response circuit based on lightwave neuromorphic signal processing could be for pilot ejection from military aircraft. Our device, which mimics the crayfish circuit using photonic technology, is sufficiently fast to be applied to defense applications in which critical decisions need to be made quickly while minimizing the probability of false alarm.

Figure 8.9 illustrates the (a) crayfish escape neuron model and (b) optical realization of the escape response for signal feature recognition. As shown in Fig. 8.9, signals from the receptors (R) are directed to the first stage of neurons—the sensory inputs (SI). Each of the SI is configured to respond to specific stimuli at the receptors. The SI integrate the stimuli and generate spikes when the inputs match the default feature. The spikes are then launched into the second stage of the neural circuit—the

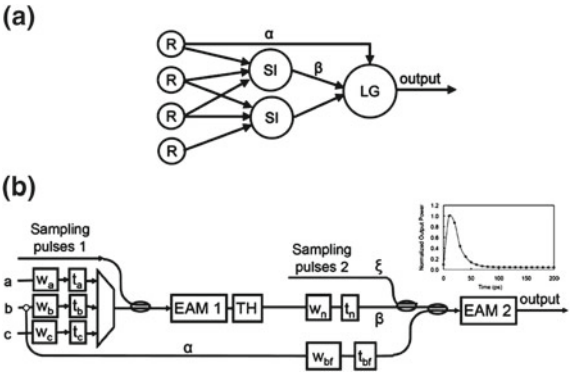


Fig. 8.9 **a** Schematic illustration of the crayfish tail-flip escape response. *R* receptors; *SI* sensory inputs; *LG* lateral giant. **b** Schematic illustration of the optical implementation of the escape response. *w* weight; *t* delay; *EAM* electro-absorption modulator; *TH* optical thresholder. Inset, measured recovery temporal profile of cross-absorption modulation in EAM

lateral giant (LG). The LG integrates the spikes from the first stage and one of the receptor signals. The neuron responds only when the signals are sufficiently close temporally and strong enough to induce a spike—an abrupt stimulus.

Our analog optical model shown in Fig. 8.9b exploits fast (subnanosecond) signal integration in electro-absorption modulators (EAM) and ultrafast (picosecond) optical thresholding in highly Ge-doped nonlinear loop mirrors (Ge-NOLM). The basic model consists of two cascaded integrators and one optical thresholder. The first integrator is configured to respond to a set of signals with specific features, while the second integrator further selects a subset of the signal from a set determined by a weighting and delay configuration and responds only when the input stimuli and the spike from the first integrator arrive within a very short time interval.

The inputs a, b, and c are weighted and delayed such that the first EAM integrator (EAM 1) is configured to spike for inputs with specific features. A train of sampling pulses is launched together with the inputs to provide a pulsed source for the EAM to spike. The spiking behavior is based on cross-absorption modulation (XAM) [27] in an EAM. That is, when the integrated input power is large enough for XAM to occur, the sampling pulses within the integration window are passed through the EAM; otherwise they are absorbed. The spike output is then thresholded at the Ge-NOLM [20] such that the output spikes are of similar height, and the undesired weak spikes are removed. The thresholded output and part of input b are launched into the second integrator as the input control through path β and α, respectively. Sampling pulses are launched to the integrator through path ξ as a spiking source. Through weighting and delaying of the inputs, spikes by the second integrator occur only for inputs with the desired features. The selection of the desired features can be reconfigured simply by adjusting the weights and delays of the inputs.

The recognition circuit detects input patterns of abc and $ab-$ having specific time intervals between the inputs as we configured. Figure 8.10 shows the experimental measurements of the signal feature recognizer. Figure 8.10a shows all eight combinations of the three inputs with specific weights and delays. We use "1" to represent the presence of input, while "0" means there is no input. Superimposed temporal profile of the input signal is shown in the inset of Fig. 8.10a. Sampling pulses with separation of $\sim$25 ps are used, as indicated by the arrows. The input signals are integrated, and the transmittance of the EAM 1 is represented by the spike pattern at the output (Fig. 8.10b). A Ge-NOLM is used to threshold the output of EAM 1 (Fig. 8.10c). The inset shows the superimposed temporal profile of the thresholded output.

When the output spikes from the first neuron arrive at the second integrator slightly after input b, i.e., within the integration interval, the second integrator will spike. By adjusting the time delay of the inputs to the EAM 2, the pattern recognizer identifies patterns abc and $ab-$ (Fig. 8.10d) or just abc (Fig. 8.10e). However, when the spikes from the first neuron arrive too late, i.e., exceed the integration time, the second integrator will not spike (Fig. 8.10f). These examples indicate that the signal feature recognizer is performing correctly and is reconfigurable through time delay adjustment.

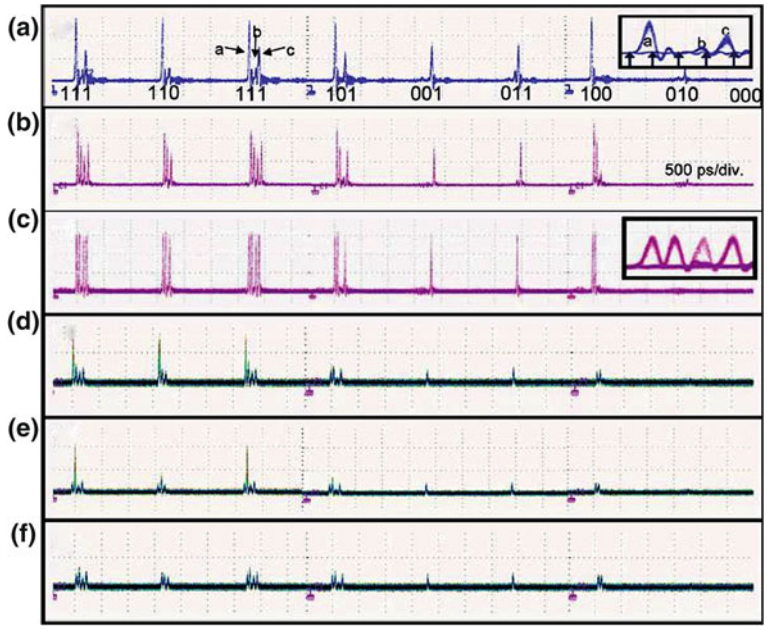


Fig. 8.10 Experimental results. **a** Optical inputs to EAM 1; **b** different spike patterns resulted from the first integrator; **c** thresheld output; **d** output spikes from EAM 2, recognizing pattern abc and $ab-$; **e** output spike from EAM 2, recognizing pattern abc only; **f** output from EAM 2, none of the input is recognized

8.7 Ultrafast Learning

Although biological neurons communicate using electro-chemical action potentials, they also possess a variety of slower, chemical regulatory processes that analyze the statistical properties of information flow. Such processes ensure network stability, maximize information efficiency, and adapt neurons to incoming signals from the environment [28, 29]. Learning as it occurs in biological neural networks can be divided into two primary categories: synaptic plasticity and intrinsic plasticity [30]. Synaptic plasticity controls the dynamics at the synapse between two communicating neurons and intrinsic plasticity controls the nonlinear transfer function of the neuron itself. Both play an important role in maximizing the mutual information between the input and output of each node and as a consequence, can perform principle component analysis (PCA) and independent component analysis (ICA) on incoming signals [31].

Our proposed photonic neural networks analogously use coherent light pulses to communicate and slower but higher density microelectronics to implement adaptive learning rules. Furthermore, successful implementation of PCA and ICA in our system will result in a robust, error-tolerant architecture that can blindly separate statistically independent unknown signals (including interference) received by antenna arrays.

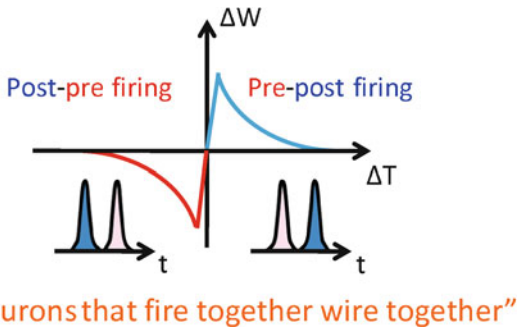


Fig. 8.11 The change in weight w_{ij} for a given edge as a function of the pre- and post- spike timings

8.7.1 Synaptic Time Dependent Plasticity

STDP is a highly parallel, linear gradient descent algorithm that tends to maximize the mutual information between the input and output of a given neuron. It operates independently on each connection and thus scales proportionally with the number of neurons times the average fan-in number. In the context of biological networks, it is an adaptive biochemical system that operates at the synapse—or gap—between communicating neurons. Its generality and simplicity allows it to be utilized for both supervised and unsupervised learning for a large variety of different learning tasks [32].

The rules governing STDP are as follows: suppose there are two neurons i, j, such that neuron i (pre-synaptic) connects to neuron j (post-synaptic) with weight w_{ij}. If neuron i fires at t_{pre} before j fires at t_{post}, STDP strengthens the weight, w_{ij}, between them. If neuron j fires before i, w_{ij} weakens. Figure 8.11 illustrates the change in weight w_{ij} as a function of the pre- and post-synaptic neuron relative spike timing difference $\Delta T = t_{\text{post}} - t_{\text{pre}}$. This plasticity curve asymmetric in ΔT is not the only kind of learning rule. Hebbian learning is an example of a symmetric kind of plasticity; however, STDP is more difficult to implement at ultrafast time scales because of the sharp discontinuity at $\Delta T = 0$. We therefore focus on STDP because it requires a photonic implementation whereas other learning rules could be implemented in slower electronic devices.

If a powerful signal travels through a network, the connections along which it travels tend to strengthen. As dictated by the STDP, since the pre-neuron i will fire before the post-neuron j, the strength between them will increase. However, misfires of the post-neuron or spike blockage of the pre-neuron will tend to decrease the corresponding weight. As a general rule, STDP emphasizes connections between neurons if they are causally related in a decision tree.

Alternatively, one can view STDP in terms of mutual information. In a given network, a neuron j can receive signals from thousands of other channels. Neurons, however, must compress that data into a single channel for their output. Since STDP strengthens the weight of causally related signals, it will attempt to minimize the

difference between the information in the input and output channels of neuron j. It thereby maximizes the *mutual information* between input and output channels.

STDP is naturally suited for unsupervised learning and cluster analysis. After feeding the network ordered input, STDP will correlate neural signals to each other, organizing the network and designate different neurons to fire for different patterns or qualities. For supervised learning, *forcing* output units at the desired values allows STDP to correlate input and output patterns. Since STDP attempts to change the network connection strengths to reflect correlation between the input and output of each node, it will automatically correlate associated examples and mold the network for the given task.

In biological networks, STDP is implemented using bio-molecular protein transmitters and receivers. This technology is noisy but scalable. Unfortunately, photonic technology cannot match the scalability of biology. Because there are $N \cdot k$ connections for a network of N neurons with mean indegree k, there is correspondingly a need for $N \cdot k$ STDP circuits to adjust each connection. This presents a scaling challege for photonic STDP because integrated photonic neural primitives themselves already approach the diffraction limit of light, and each may have many inputs.

8.7.2 Intrinsic Plasticity

Intrinsic Plasticity (IP) describes a set of adaptive algorithms that regulate the internal dynamics of neurons rather than the synapses connecting them [33]. IP operates within the neuron after STDP has applied the weights. Unlike STDP, it refers to a class of adaptations instead of a unified learning rule. Researchers posit that it tends to maximize the mutual information between a neuron's input and output channels, like STDP [34]. Since IP controls spiking dynamics rather than connection strengths, it scales with the number of neurons, N and does not present a significant architectural challenge. Photonic neurons exhibit changes in their dynamics with changes in the current injected into the semiconductor, providing this as a mechanism for photonic IP. The combination of IP algorithms with STDP encourages network stability and allows for a higher diversity of applications, including (ICA) [31].

8.7.3 Principal Component Analysis

PCA seeks to find the eigenvector(s) of a high-dimensional space of data along the direction of maximal variance. This assures that an orthogonal projection of the data into a desired subspace retains the maximum information from the original data set as possible. Spiking neurons can extract the principle components of incoming data using a combination of STDP, synaptic scaling (SS), and lateral inhibition. In this kind of circuit, signals are simply encoded as spike rates, where the amplitude of

an incoming signal modulates the rate of fixed amplitude spikes. Spike-based rate coding is more robust against analog noise than direct analog modulation.

Though neural weights are controlled by slow electronic integrators, a photonic STDP circuit (see Fig. 8.13a) is needed to distinguish pulse arrival times with picosecond resolution. The weight learning rule for STDP is given by a piecewise exponential curve in the spike relative arrival time domain (see Fig. 8.13b).

Synaptic scaling normalizes the sum of weights across all the inputs for a given neuron, preventing STDP from destabilizing the weights and introducing competition between weights. This normalization is implemented electronically. Lateral inhibition between neurons in the output layer decouples signals from each neuron, forcing them to avoid correlated outputs. This effect is required for extraction of principal components other than the first.

The PCA algorithm takes N inputs encoded in spike rates along N channels. Neurons output a projection of this data onto an initial vector (the *weight vector*) determined by the strength of the connections to that neuron. STDP is set to operate in a simple Hebbian learning scheme in which a positive change in weight depends multiplicatively on both the input and output firing rates. As each STDP circuit operates in parallel along each of the N dimensions, the net effect is an adaptive adjustment of the weight vector towards the principle component eigenvector (see Fig. 8.12). Subsequent principle components can also be extracted by subtracting higher principle components from other neurons using lateral inhibition, forcing them to extract information from what remains.

8.7.4 Independent Component Analysis

Although PCA is useful for a variety of tasks, it cannot perform blind source separation if the signals are not mixed orthogonally. ICA, on the other hand, can separate

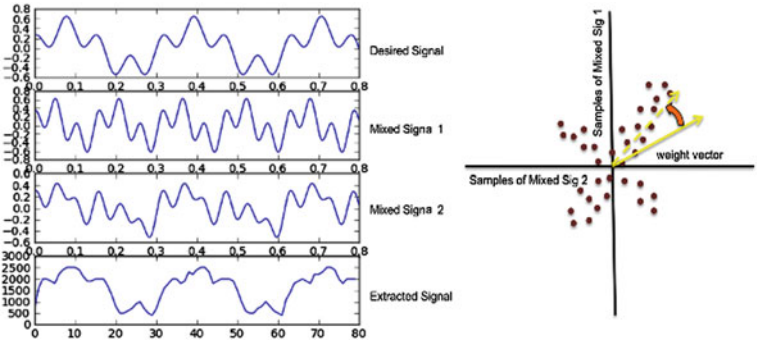


Fig. 8.12 An example of signal demixing with PCA. The top signal is mixed with an interfering signal in two channels, leading to the second two signals. Simulations with STDP and synaptic scaling show that a single neuron can extract the original signal from the mixed signals using PCA

multivariate data into its independent subcomponents for an arbitrary unknown mixing matrix. ICA operates effectively under the conditions that the numbers of inputs (or antennas) are at least as great as the number of original signals and the signals have non-Gaussian distributions. Spiking neurons can perform ICA with the addition of IP which can be implemented with photodetectors to measure activity and electronic circuits to implement adaptation.

The ICA algorithm operates similarly to the PCA algorithm with two modifications: (1) STDP can decrease the strength of the connection for inputs that initiate low outputs, and (2) IP can change the slope of the neuron's transfer function based on the current activity of the neuron. The interaction of STDP with IP results in a learning scheme that favors highly kurtotic dimensions, allowing the extraction of non-orthogonal independent components [31]. Like in PCA, subsequent independent components can be extracted using lateral inhibition.

8.7.5 Photonic STDP

Artificial STDP has been explored in VLSI electronics [35], and more recently, has been proposed [36] and demonstrated [37] in memristive nanodevices. Some ongoing projects in microelectronics seek to develop hardware platforms based on this technology [3, 4]. The analog nature and resilience to noise of neuromorphic processing naturally complements the high variability and failure rates of nanodevices [6], enabling the possibility of densely connected adaptive spiking networks [38]. STDP was first explored in the optical domain by [39].

The optical STDP circuit is illustrated in Fig. 8.13. This implementation uses two slow integrators and optical summing to create an exponential-like response function. The resulting response is then incident on a photodetector, which regulates an electronic circuit to control the weight between two neurons. The response function shape can be dynamically adjusted with various control parameters (Fig. 8.14). This photonic STDP design can be integrated, but the exceptional need for up to $N \cdot k$ independent units limits the scalability of an overall system. Analogous STDP devices proposed in electronics such as memristors in nano-crossbar arrays [6] overcome this scaling challenge with extremely small nano-devices (on the order of nanometers). Novel implementations of STDP based on electronic-optical, photonic crystal [40], or plasmonic [41] technologies may become important to support scaling of complete adaptability. The requirement of every connection to adapt without supervision can also be relaxed, for example, by organizing the system as a liquid state machine (LSM) or in other reservoir architecture [42].

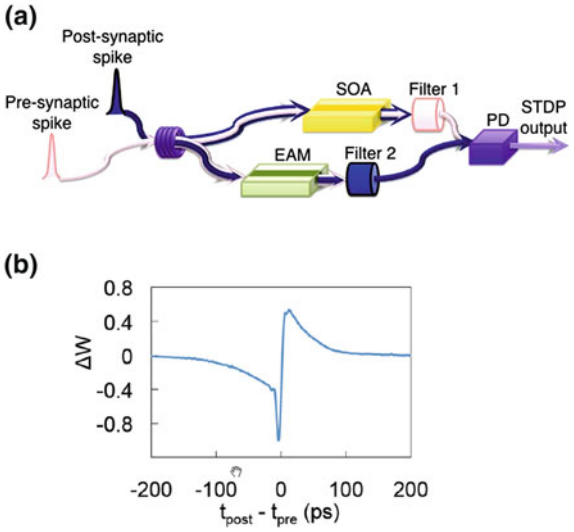


Fig. 8.13 **a** Photonic circuit diagram of STDP and **b** experimental function of output power vs. spike interval. Pre and post synaptic inputs cross phase modulate one another, with time constants determined by the SOA and EAM to give the characteristic STDP curve

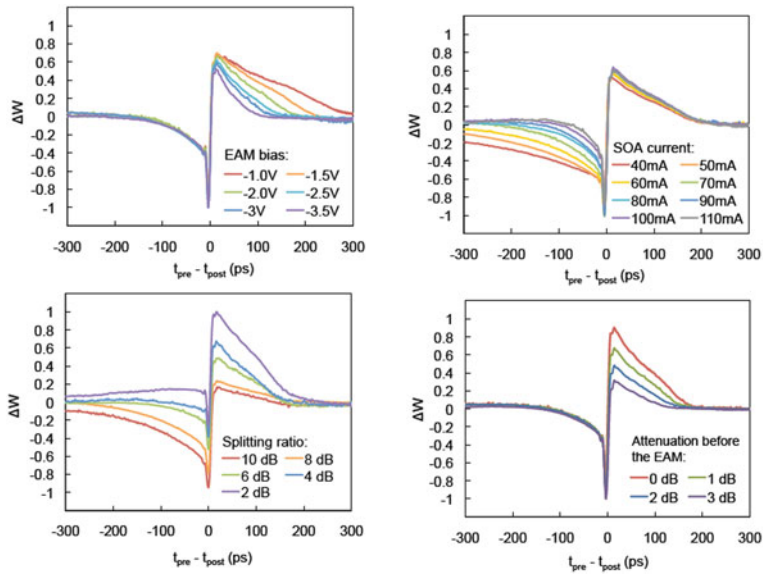


Fig. 8.14 Adjustable parameters in both the EAM and SOA can dynamically change the properties of the STDP circuit

8.8 Excitable Laser Neuron

The fiber-based bench-top photonic neuron provided a starting basis for investigations into photonic implementations of spiking neuromorphic devices. The primary reason for moving beyond it is integrability. Photonic integration, much like electronic integration, enables much greater scalability, which is required to access a much more interesting and rich computational repertoire based on large groups of neurons. We have found that laser dynamics constrained to a dynamically excitable regime can also emulate the LIF model, in many ways more accurately and unambiguously than the initial bench-top device did [43, 44]. Experiments with a fiber-based prototype laser with a graphene SA have confirmed the ability of this laser model to achieve excitable dynamics [45, 46]. Additional advantages of integration include very low power operation, hardware cost reduction, and added robustness to environmental fluctuations. In this section, we reveal the analogy between the LIF model and a modified excitable laser model.

Our starting point is a set of dimensionless equations describing SA lasers that can generalize to a variety of different laser systems, including passively Q-switched microchip lasers [47], distributed BRAGG reflector lasers [48], and vertical cavity surface emitting lasers (VCSELs) [49]. The Yamada model describes the behavior of lasers with independent gain and saturable absorber (SA) sections with an approximately constant intensity profile across the cavity [50], as illustrated in Fig. 8.15. We assume that the SA has a very short relaxation time on the order of the cavity intensity, which can easily be achieved either through doping or special material properties. The dynamics now operate such that the gain is a slow variable while the intensity and loss are both fast. This three-dimensional dynamical system can be described with the following equations:

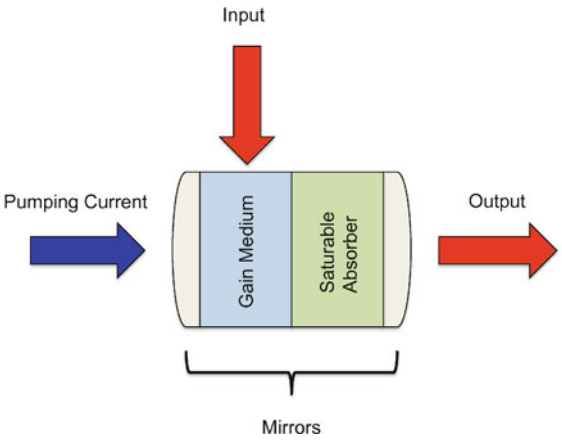


Fig. 8.15 A simple schematic of a SA laser. The device is composed of (i) a gain section, (ii) a saturable absorber, and (iii) mirrors for cavity feedback. In the LIF excitable model, inputs selectively perturb the gain optically or electrically

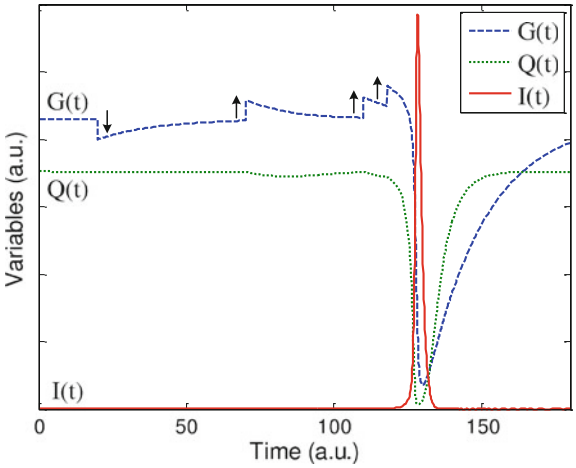


Fig. 8.16 Simulation results of an SA laser behaving as a LIF neuron. *Arrows* indicate excitatory pulses and inhibitory pulses that change the gain by some amount $\triangle G$. Enough excitatory input causes the system to enter fast dynamics in which a spike is generated, followed by the fast recover of the absorption $Q(t)$ and the slow recover of the gain $G(t)$. Variables were rescaled to fit within the desired range. Values used: $A = 4.3$; $B = 3.52$; $a = 1.8$; $\gamma_G = 0.05$; $\gamma_L, \gamma_I >> 0.05$

$$\dot{G(t)} = \gamma_G \left[A - G(t) - G(t)I(t)\right] \tag{8.4}$$

$$\dot{Q(t)} = \gamma_Q \left[B - Q(t) - aQ(t)I(t)\right] \tag{8.5}$$

$$\dot{I(t)} = \gamma_I \left[G(t) - Q(t) - 1\right] I(t) + \epsilon f(G) \tag{8.6}$$

where $G(t)$ models the gain, $Q(t)$ is the absorption, $I(t)$ is the laser intensity, A is the bias current of the gain, B is the level of absorption, a describes the differential absorption relative to the differential gain, γ_G is the relaxation rate of the gain, γ_Q is the relaxation rate of the absorber, γ_I is the inverse photon lifetime, and $\epsilon f(G)$ represents the small contributions to the intensity made by spontaneous emission (noise term) where ϵ is very small. Although conventionally γ_I is set to 1, we include the reverse photon lifetime γ_I for clarity.

We further assume that inputs to the system cause perturbations to the gain $G(t)$ only. Pulses—from other excitable lasers, for example—will induce a change $\triangle G$ as illustrated by the arrows in Fig. 8.16 and analog inputs will modulate $G(t)$ continuously. This injection can be achieved either via optical pulses that selectively modulate the gain medium or through electrical current injection.

8.8.1 Before Pulse Formation

Since the loss $Q(t)$ and the intensity $I(t)$, are fast, they will quickly settle to their equilibrium values. On slower time scales, our system behaves as:

$$\dot{G(t)} = \gamma_G \left[A - G(t) - G(t)I(t)\right] + \theta(t) \tag{8.7}$$

$$Q(t) = Q_{\text{eq}} \tag{8.8}$$

$$I(t) = I_{\text{eq}} \tag{8.9}$$

with $\theta(t)$ representing possible inputs, and the equilibrium values $Q_{eq} = B$ and $I_{eq} = \epsilon f(G)/\left[\gamma_I \left(1 - G(t) + Q(t)\right)\right]$. Since ϵ is quite small, $I_{\text{eq}} \approx 0$. With zero intensity in the cavity, the $G(t)$ and $Q(t)$ variables are dynamically decoupled. The result is that if inputs are incident on the gain, they will only perturb $G(t)$ unless $I(t)$ becomes sufficiently large to couple the dynamics together.

If $I(t)$ increases, the slow dynamics will break. $I(t)$ will become unstable when $G(t) - Q(t) - 1 > 0$ because $\dot{I}(t) \approx \gamma_I \left[G(t) - Q(t) - 1\right] I(t)$. Given our perturbations to $G(t)$, we can define a threshold condition:

$$G_{\text{thresh}} = Q + 1 = B + 1 \text{ (at equilibrium)} \tag{8.10}$$

above which fast dynamics will take effect. This occurs after the third excitatory pulse in Fig. 8.16.

8.8.2 Pulse Generation

Perturbations that cause $G(t) > G_{\text{thresh}}$ will result in the release of a short pulse. Once $I(t)$ is lifted above the attractor $I = 0$, $I(t)$ will increase exponentially. This results in the saturation of $Q(t)$ on a very fast time scale until $Q = 0$, followed by the slightly slower depletion of the gain, $G(t)$. Once $G(t) - Q(t) - 1 < 0$, that is, $G(t) < 1$ following saturation, $I(t)$ will hit its peak intensity I_{max}, followed by a fast decay on the order of $1/\gamma_I$ in time. $I(t)$ will eventually reach $I \approx 0$ as it further depletes the gain to a final value G_{reset}, which—with a large enough intensity—is often close to the transparency level $G_{\text{reset}} \approx 0$.

A given pulse derives its energy from excited carriers in the cavity. The total energy of the pulse is $E_{pulse} = Nh\nu$, where N is the number of excited carriers that have been depleted and $h\nu$ is the energy of a single photon at the lasing frequency. Because the gain is proportional to the inversion population, N must be proportional to the amount that the gain $G(t)$ has depleted during the formation of a pulse. Thus, if G_{fire} is the gain that causes the release of a pulse, we can expect that an output pulse will take the approximate form:

$$P_{\text{out}} = E_{\text{pulse}} \cdot \delta(t - \tau_f) \tag{8.11}$$

$$E_{pulse} \propto G_{\text{fire}} - G_{\text{reset}} \tag{8.12}$$

where τ_f is the time at which a pulse is triggered to fire and $\delta(t)$ is a delta function. One of the properties of spike-encoded channels is that spike energies are encoded digitally. Spikes must have a constant amplitude every iteration, a

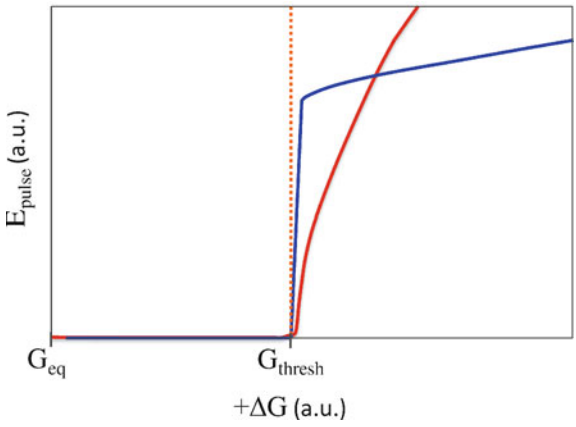


Fig. 8.17 Neuron threshold functions. The *red* and *blue curves* are simulated, normalized transfer functions for a single input spike when the neuron is operated far from and close to the threshold, respectively. Setting G_{eq} close to G_{thresh} reduces the required perturbation $\triangle G$ to initiate a pulse and thereby minimizes the impact it has on the resulting output pulse, leading to the flatter one level region on the blue curve

characteristic property of the *all-or-nothing* response shared by biological neurons. We can normalize our output pulses if we set our system to operate close to threshold $G_{\text{thresh}} - G_{eq} \ll G_{\text{thresh}}$. Since the threshold is effectively lowered, the size of input perturbations $\triangle G$ must be scaled smaller. This implies $G_{fire} \approx G_{\text{thresh}}$, which helps in suppressing variations in the output pulse amplitude by reducing the input perturbation to the system. This leads to a step-function like response, as illustrated in Fig. 8.17, which is the desired behavior.

After a pulse is released, $I(t) \to 0$ and $Q(t)$ will quickly recover to Q_{eq}. The fast dynamics will give way to slower dynamics, in which $G(t)$ will slowly creep from G_{reset} to G_{eq}. The fast dynamics of $Q(t)$ assure that the threshold $G_{\text{thresh}} = 1 + Q(t)$ recovers quickly after a pulse is generated, preventing partial pulse release during the recovery period. In addition, the laser will experience a *relative refractory period* in which it is difficult—but not impossible—to fire another pulse.

8.8.3 LIF Analogy

If we assume the fast dynamics are nearly instantaneous, we can compress the behavior of our system into the following set of equations and conditions:

$$\frac{dG(t)}{dt} = -\gamma_G(G(t) - A) + \theta(t); \tag{8.13}$$

$$\text{if } G(t) > G_{\text{thresh}} \text{ then} \tag{8.14}$$

$$\text{release a pulse, and set } G(t) \to G_{\text{reset}}.$$

where $\theta(t)$ represent input perturbations. This behavior can be seen qualitatively in Fig. 8.16. The conditional statements account for the fast dynamics of the system that occur on times scales of order $1/\gamma_I$, and other various assumptions—including the

fast $Q(t)$ variable and operation close to threshold—assure that G_{thresh}, G_{reset} and the pulse amplitude E_{pulse} remain constant. If we compare this to the LIF model, or (8.1):

$$C_m \frac{dV_m(t)}{dt} = -\frac{1}{R_m}(V_m(t) - V_L) + I_{app}(t);$$
$$\text{if} V_m(t) > V_{\text{thresh}} \text{ then}$$
$$\text{release a spike and set } V_m(t) \rightarrow V_{\text{reset}}.$$

the analogy between the equations becomes clear. Setting the variables $\gamma_G = 1/R_m C_m$, $A = V_L$, $\theta(t) = I_{app}(t)/R_m C_m$, and $G(t) = V_m(t)$ shows their algebraic equivalence. Thus, the gain of the laser $G(t)$ can be thought of as a virtual *membrane voltage*, the input current A as a virtual *leakage voltage*, etc. There is a key difference, however—both dynamical systems operate on vastly different time scales. Whereas biological neurons have time constants $\tau_m = C_m R_m$ on order of milliseconds, carrier lifetimes of laser gain sections are typically in the *ns* range and can go down to *ps*.

8.8.4 Excitable VCSELs

Although the excitable model is generalizable to a variety of different laser types, vertical cavity surface emitting lasers (VCSEL) are a particularly attractive candidate for our computational primitive as they occupy small footprints, can be fabricated in large arrays allowing for massive scalability, and use low powers [51]. An excitable VCSEL with an intra-cavity SA that operates using the same rate equation model described above has already been experimentally realized [52]. In addition, the technology is amenable to a variety of different interconnect schemes: VCSELs can send signals upward and form 3D interconnects [53], can emit downward into an interconnection layer via grating couplers [54] or connect monolithically through intra-cavity holographic gratings [55].

A schematic of our VCSEL structure, which includes an intra-cavity SA, is illustrated in Fig. 8.18. To simulate the device, we use a typical two-section rate equation model such as the one described in [49]:

$$\frac{dN_{ph}}{dt} = \Gamma_a g_a (n_a - n_{0a}) N_{ph} + \Gamma_s g_s (n_s - n_{0s}) N_{ph} - \frac{N_{ph}}{\tau_{\text{ph}}} + V_a \beta B_r n_a^2 \tag{8.15}$$

$$\frac{dn_a}{dt} = -\Gamma_a g_a (n_a - n_{0a}) \frac{(N_{ph} - \phi(t))}{V_a} - \frac{n_a}{\tau_a} + \frac{I_a}{eV_a} \tag{8.16}$$

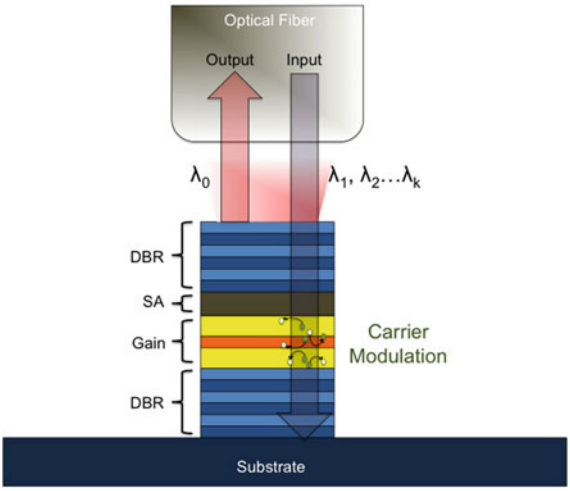


Fig. 8.18 A schematic diagram of an excitable VCSEL interfacing with a fiber leading to the network. In this configuration, inputs $\lambda_1, \lambda_2 \ldots \lambda_n$ selectively modulate the gain section. Various frequencies lie on different parts of the gain spectrum, leading to wavelength-dependent excitatory or inhibitory responses. The weights and delays are applied by amplifiers and delay lines within the fiber network. If excited, a pulse at wavelength λ_0 is emitted upward and transmitted other excitable lasers

$$\frac{dn_s}{dt} = -\Gamma_s g_s (n_s - n_{0s}) \frac{N_{ph}}{V_s} - \frac{n_s}{\tau_s} + \frac{I_s}{eV_s} \tag{8.17}$$

where $N_{ph}(t)$ is the total number of photons in the cavity, $n_a(t)$ is the number of carriers in the gain region, and $n_s(t)$ is the number of carriers in the absorber. Subscripts a and s identify the active and absorber regions, respectively. The remaining device parameters are summarized in Table 8.3. We add an additional input term $\phi(t)$ to account for optical inputs selectively coupled into the gain and an SA current injection term I_s/eV_s to allow for an adjustable threshold.

These equations are analogous to the dimensionless set of (8.4) provided that the following coordinate transformations are made:

$$G(\tilde{t}) = \tau_{ph}\Gamma_a g_a (n_a(t) - n_{0a}), \quad I(\tilde{t}) = \frac{\tau_a \Gamma_a g_a}{V_a} N_{ph}(t)$$
$$Q(\tilde{t}) = \tau_{ph}\Gamma_s g_s (n_{0s} - n_s(t)), \quad \tilde{t} = \frac{t}{\tau_{ph}}$$

where differentiation is now with respect to $\tilde{t}$ rather than t. The dimensionless parameters are now

$$\gamma_G = \frac{\tau_{ph}}{\tau_a}, \quad A = \tau_a \tau_{ph} \Gamma_a g_a \left[\frac{I_a}{eV_a} - \frac{n_{0a}}{\tau_a}\right],$$
$$\gamma_Q = \frac{\tau_{ph}}{\tau_s}, \quad B = \tau_s \tau_{ph} \Gamma_s g_s \left[\frac{n_{0s}}{\tau_s} - \frac{I_s}{eV_s}\right],$$
$$\gamma_I = 1, \qquad a = \frac{\tau_s \Gamma_s g_s V_a}{\tau_a \Gamma_a g_a V_s},$$

Table 8.3 VCSEL-SA excitable laser parameters [49, 56–58]

Parameter	Description	Value
λ	Lasing wavelength	850 nm
V_a	Active region cavity volume	$2.4\times10^{-18}\,\mathrm{m}^3$
V_s	SA region cavity volume	$2.4\times10^{-18}\,\mathrm{m}^3$
Γ_a	Active region confinement factor	0.06
Γ_s	SA region confinement factor	0.05
τ_a	Active region carrier lifetime	1 ns
τ_s	SA region carrier lifetime	100 ps
τ_p	Photon lifetime	4.8 ps
g_a	Active region differential gain/loss	$2.9\times10^{-12}\,\mathrm{m}^3\mathrm{s}^{-1}$
g_s	SA region differential gain/loss	$14.5\times10^{-12}\,\mathrm{m}^3\mathrm{s}^{-1}$
n_{0a}	Active region transparency carrier density	$1.1\times10^{24}\,\mathrm{m}^{-3}$
n_{0s}	SA region transparency carrier density	$0.89\times10^{24}\,\mathrm{m}^{-3}$
B_r	Bimolecular recombination term	$10\times10^{-16}\,\mathrm{m}^3\mathrm{s}^{-1}$
β	Spontaneous emission coupling factor	1×10^{-4}
η_c	Output power coupling coefficient	0.4

$$\epsilon f(G) = \tau_a \tau_{ph} \Gamma_a g_a \beta B_r \left[\frac{G}{\tau_{ph} \Gamma_a g_a} + n_{0a} \right]^2$$

For the simulation, we set the input currents to $I_a = 2\,\mathrm{mA}$ and $I_s = 0\,\mathrm{mA}$ for the gain and absorber regions, respectively. The output power is proportional to the photon number N_{ph} inside the cavity via the following formula:

$$P_{\mathrm{out}}(t) \approx \frac{\eta_c \Gamma_a}{\tau_p} \frac{hc}{\lambda} N_{ph}(t) \tag{8.18}$$

in which η_c is the output power coupling coefficient, c the speed of light, and hc/λ is the energy of a single photon at wavelength λ. We assume the structure is grown on a typical GaAs-based substrate and emits at a wavelength of 850 nm.

Using the parameters described above, we simulated the device with optical injection into the gain as shown in Fig. 8.19. Input perturbations that cause gain depletion or enhancement—represented by positive and negative input pulses—modulate the carrier concentration inside the gain section. Enough excitation eventually causes the laser to enter fast dynamics and fire a pulse. This behavior matches the behavior of an LIF neuron described in Sect. 8.8.3.

Our simulation effectively shows that an excitable, LIF neuron is physically realizable in a VCSEL-SA cavity structure. The carrier lifetime of the gain is on the order of 1 ns, which as we have shown in Sect. 8.8.3 is analogous to the $R_m C_m$ time constant of a biological neuron—typically on the order of 10 ms. Thus, our device already exhibits speeds that are 10 million times faster than a biological equivalent. Lifetimes could go as short as ps, making the factor speed increase between biology and photonics up to a billion.

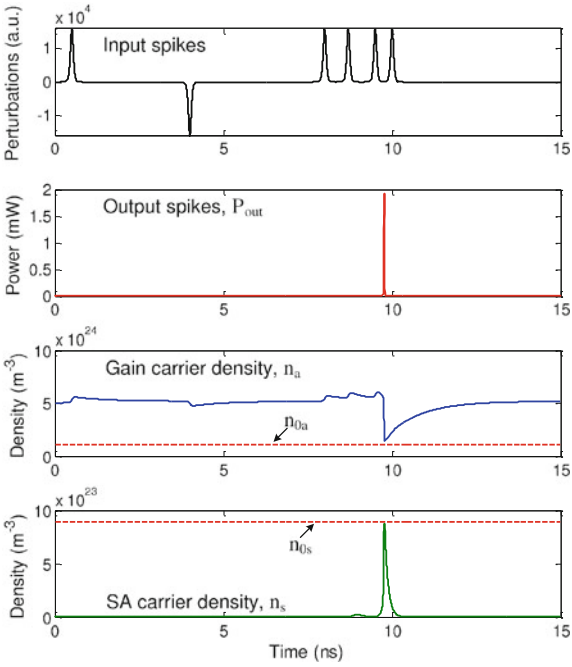


Fig. 8.19 Simulation of an excitable, LIF excitable VCSEL with realistic parameters. Inputs (*top*) selectively modulate the carrier concentration in the gain section (*middle*). Enough excitation leads to the saturation of the absorber to transparency (*bottom*) and the release of a pulse, followed by a relative refractory period while the pump current recovers the carrier concentration back to its equilibrium value

8.8.5 Other Spiking Photonic Devices

Since the early bench-top model of a spiking photonic neuron suggested an ultrafast cognitive computing domain, several other approaches have arisen to develop scalable, integrated photonic spike processing devices. One device designed and demonstrated by [59, 60] uses the nonlinear dynamics of polarization switching (PS) in a birefringent VCSEL to emulate the behavior of specialized kinds of neurons. Characteristic behaviors of the resonate-and-fire neural model [61] (as opposed to the above integrate-and-fire model) including tonic spiking, rebound spiking, and subthreshold oscillations have been observed. The state of the VCSEL depends on the lasing power of competetive orthogonally polarized modes, so the direction of input fluence (excitatory or inhibitory) is determined by the injected light polarization and wavelength.

Semiconductor ring lasers (SRL) have also been shown to exhibit dynamic excitability [62]. They are investigated for application as a computational primitive in [63]. Excitable bifurcations in SRLs can arise from weakly broken $\mathbb{Z}_2$ symmetry between counterpropagating normal modes of the ring resonator, where $\mathbb{Z}_2$ refers to the complex two-dimensional phase space that describes the SRL state. The direction of input fluence on SRL state is strongly modulated by the optical phase of the circulating and input fields due to this complex phase space characteristic of optical degrees of freedom. These neuron-like behaviors currently under investigation are

so far all based on different constrained regimes of the dynamically rich laser equations. They differ primarily in their physical representation of somatic integration variables, spiking signals, and correspondingly the mechanism of influence of one on another.

8.9 Cortical Spike Algorithms: Small-Circuit Demos

This section describes implementation of biologically-inspired circuits with the excitable laser computational primitive. These circuits are rudimentary, but fundamental exemplars of three spike processing functions: multistable operation, synfire processing [64], and spatio-temporal pattern recognition [65].

We stipulate a mechanism for optical outputs of excitable lasers to selectively modulate the gain of others through both excitatory (gain enhancement) and inhibitory (gain depletion) pulses, as illustrated in Fig. 8.18. Selective coupling into the gain can be achieved by positioning the gain and saturable absorber regions to interact only with specific optical frequencies as experimentally demonstrated in [52]. Excitation and inhibition can be achieved via the gain section's frequency dependent absorption spectrum—different frequencies can induce gain enhancement or depletion. This phenomenon been experimentally demonstrated in semiconductor optical amplifiers (SOAs) [19] and could generalize to laser gain sections if the cavity modes are accounted for. Alternatives to these proposed solutions include photodetectors with short electrical connections and injection into an extended gain region in which excited carriers are swept into the cavity via carrier transport mechanisms.

A network of excitable lasers connected via weights and delays—consistent with the model described in Sect. 8.4—can be described as a delayed differential equation (DDE) of the form:

$$\frac{d}{dt}\boldsymbol{x}(t) = f(\boldsymbol{x}(t), \boldsymbol{x}(t-\tau_1), \boldsymbol{x}(t-\tau_2) \cdots \boldsymbol{x}(t-\tau_n)) \tag{8.19}$$

where the vector $\boldsymbol{x}(t)$ contains all the state variable associated with the system. The output to our system is simply the output power, $\boldsymbol{P}_{\mathbf{out}}(t)$,[2] while the input is a set of weighted and delayed outputs from the network, $\sigma(t) = \sum_k W_k \boldsymbol{P}_{\mathbf{out}}(t-\tau_k)$. We can construct weight and delay matrices W, D such that the W_{ij} element of W represents the strength of the connection between excitable lasers i, j and the D_{ij} element of D represents the delay between lasers i, j. If we recast (8.15) in a vector form, we can formulate our system in (8.19) given that the input function vector $\phi(t)$, is

$$\phi(t) = \Omega \boldsymbol{\Theta}(t) \tag{8.20}$$

[2] We absorb the attenuation or amplification the pulse experiences en route to its destination along with the responsivity of the perturbation to the incident pulse into a single weight parameter W_{ij}.

where we create a sparse matrix Ω containing information for both W and D, and a vector $\boldsymbol{\Theta}(t)$ that contains all the past outputs from the system during unique delays $U = [\tau_1,\ \tau_2,\ \tau_3\ \cdots\ \tau_n]$:

$$\Omega = [W_0\ W_1\ W_2 \cdots W_n] \tag{8.21}$$

$$\boldsymbol{\Theta}(t) = \begin{pmatrix} \boldsymbol{P}_{\mathbf{out}}(t) \\ \boldsymbol{P}_{\mathbf{out}}(t-\tau_1) \\ \boldsymbol{P}_{\mathbf{out}}(t-\tau_2) \\ \vdots \\ \boldsymbol{P}_{\mathbf{out}}(t-\tau_n) \end{pmatrix} \tag{8.22}$$

W_k describes a sparse matrix of weights associated with the delay in element k of the unique delay vector U. To simulate various system configurations, we used Runge-Kutta methods iteratively within a standard DDE solver in MATLAB. This formulation allows the simulation of arbitrary networks of excitable lasers which we used for several cortical-inspired spike processing circuits described below.

8.9.1 Multistable System

Multistability represents a crucial property of dynamical systems and arises out of the formation of hysteretic attractors. This phenomenon plays an important role in the formation of memory in processing systems. Here, we describe a network of two interconnected excitable lasers, each with two incoming connections and identical weights and delays, as illustrated in Fig. 8.20. The system is recursive rather than feedforward, which results in a settable dynamic memory element similar to a digital flip-flop.

Results for the two laser multistable system are shown in Fig. 8.20. The network is composed of two lasers, interconnected via optical connections with a delay of 1 ns. An excitatory pulse travels to the first unit at $t = 5$ ns, initiating the system to settle to a new attractor. The units fire pulses repetitively at fixed intervals before being deactivated by a precisely timed inhibitory pulse at $t = 24$ ns. It is worth noting that the system is also capable of stabilizing to other states, including those with multiple pulses or different pulse intervals. It therefore acts as a kind of optical pattern buffer over longer time scales. Ultimately, this circuit represents a test of the network's ability to handle recursive feedback. In addition, the stability of the system implies that a network is cascadable since a self-referent connection is isomorphic to an infinite chain of identical lasers with identical weights W between every node. Because this system successfully maintains the stability of self-pulsations, processing networks of excitable VCSELs are theoretically capable of cascadibility and information retention during computations.

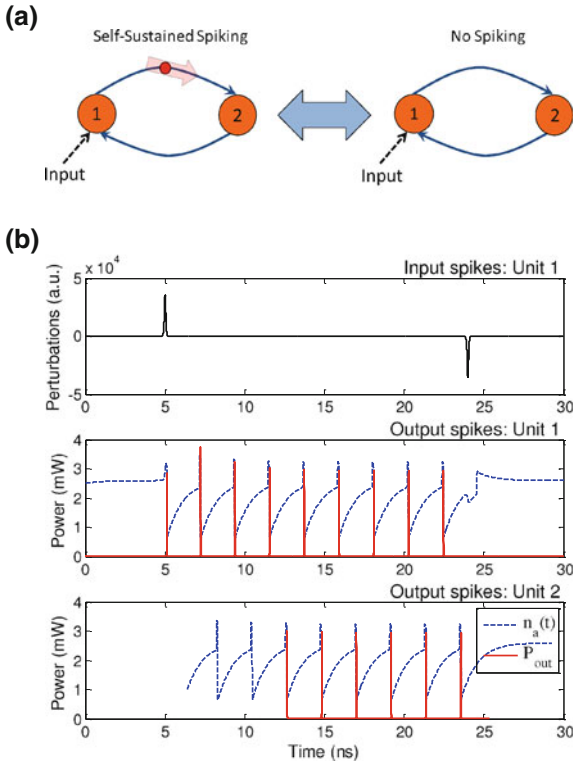


Fig. 8.20 **a** Bistability schematic. In this configuration, two lasers are connected symmetrically to each other. **b** A simulation of a two laser system exhibiting bistability with connection delays of 1 ns. The input perturbations to unit one are plotted, followed by the output powers of units 1 and 2 , which include scaled version of the carrier concentrations of their gain sections as the *dotted blue lines*. Excitatory pulses are represented by positive perturbations while inhibitory pulses are represented by negative perturbations. An excitatory input excites the first unit, causing a pulse to be passed back and forth between the nodes. A precisely timed inhibitory pulse terminates the sequence

8.9.2 Synfire Chain

Synfire chains have been proposed by Abeles [66] as a model of cortical function. A synfire chain is essentially a feedforward network of neurons with many layers (or pools). Each neuron in one pool feeds many excitatory connections to neurons in the next pool, and each neuron in the receiving pool is excited by many neurons in the previous pool, so that a wave of activity can propagate from pool to pool in the chain. It has been postulated that such a wave corresponds to an elementary cognitive event [67].

Synfire chains are a rudimentary circuit for *population encoding*, which reduces the rate of jitter accumulation when sending, receiving, or storing a spatio-temporal bit pattern of spikes [64]. Population encoding occurs when multiple copies of a signal are transmitted along W uncorrelated channels (W for chain *width*). When these copies arrive and recombine in subsequent integrator units, statistically uncorrelated jitter and amplitude noise is averaged resulting in a noise factor that is less than the original by a factor of $W^{-1/2}$. One of the key features of a hybrid analog-digital system such as an SNN is that many analog nodes can process in a distributed and redundant way to reduce noise accumulation. Recruiting a higher number of neurons

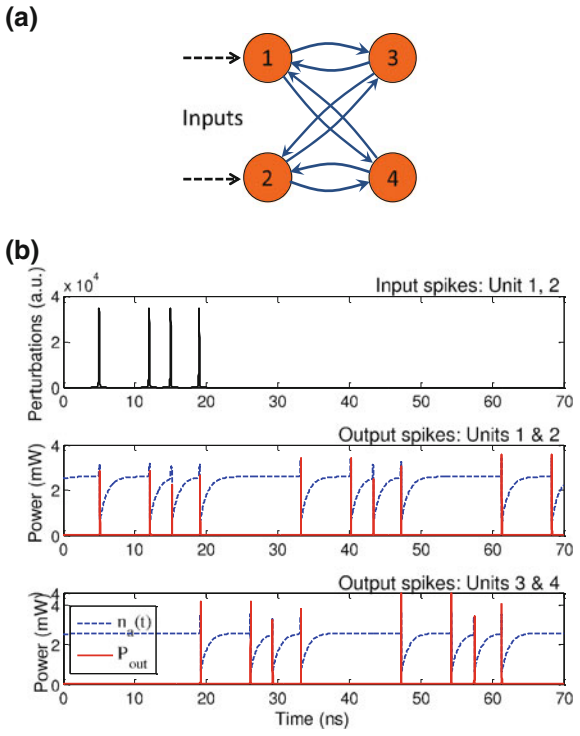


Fig. 8.21 **a** Synfire schematic. In this configuration, two groups of lasers are connected symmetrically two each other. **b** Simulation of a four laser circuit forming synfire chains with connection delays of 14 ns. The input perturbations to units 1, 2 are plotted over time, followed by the output powers of units 1–4 with the scaled carrier concentrations of their gain sections as the dotted blue lines. A characteristic spike pattern is repeatedly passed back and forth between the *left* and *right* set of nodes

to accomplish the same computation is an effective and simple way of reducing spike error rates.

Figure 8.21a shows a demonstration of a simple four-laser synfire chain with feedback connections. The chain is simply a two unit expansion of each node in the multistability circuit from Fig. 8.20a. Like the multistability circuit, recursion allows the synfire chain to possess hysteric properties; however, the use of two lasers for each logical node provides processing redundancy and increases reliability. Once the spike pattern is input into the system as excitatory inputs injected simultaneously into the first two lasers, it is continuously passed back and forth between each set of two nodes. The spatio-temporal bit pattern persists after several iterations and is thereby stored in the network as depicted in Fig. 8.21b.

8.9.3 Spatio-Temporal Pattern Recognition Circuit

The concept of polychrony, proposed by Izhikevich [65] is defined as an event relationship that is precisely time-locked to firing patterns but not necessarily synchronous. Polychronization presents a minimal spiking network that consists of cortical spiking neurons with axonal delays and synaptic time dependent plasticity (STDP),

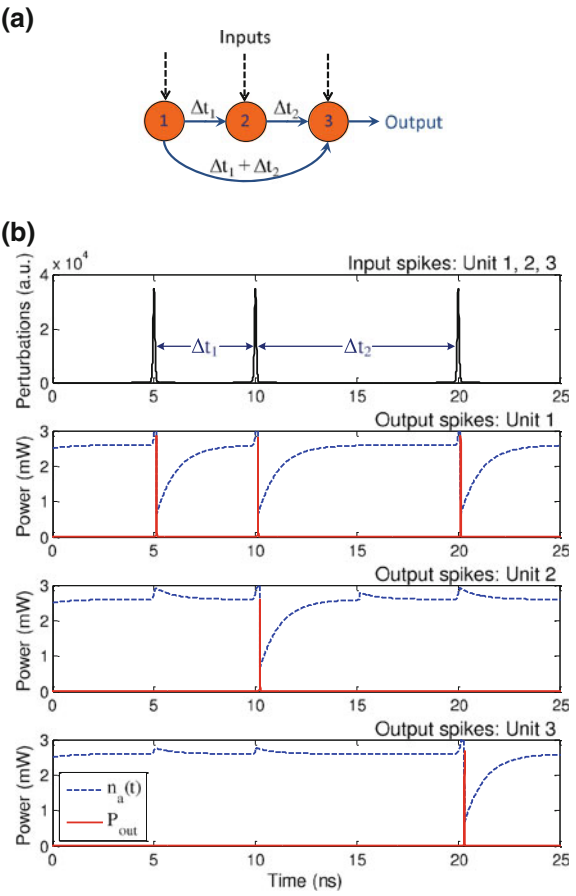


Fig. 8.22 **a** Schematic of a three-laser circuit that can recognize specific spatio-temporal bit patterns. **b** A simulation of a spatio-temporal recognition circuit with $\triangle t_1 = 5\,\text{ns}$ and $\triangle t_2 = 10\,\text{ns}$. The input perturbation to unit 1 is plotted, along with the output powers of units 1–3 with the scaled carrier concentrations of their gain sections as the *dotted blue lines*. The third neuron fires during the triplet spike pattern with time delays $\triangle t_1$ and $\triangle t_2$ between spikes

an important learning rule for spike-encoded neurons. As a result of the interplay between the delays and STDP, spiking neurons spontaneously self-organize into groups and generate patterns of stereotypical polychronous activity.

One of the key properties of polychronization is the ability to perform *delay logic* to perform spatio-temporal pattern recognition. As shown in Fig. 8.22a, we construct a simple three unit pattern recognition circuit our of excitable lasers with carefully tuned delay lines, where each subsequent neuron in the chain requires stronger perturbations to fire. The resulting simulation is shown in Fig. 8.22b. Three excitatory inputs separated sequentially by $\Delta t_1 = 5$ ns and $\Delta t_2 = 10$ ns are incident on all three units. The third is configured only to fire if it receives an input pulse and pulses from the other two simultaneously. The system therefore only reacts to a specific spatio-temporal bit pattern.

Although this circuit is simple, the ability to perform temporal logic implies that excitable, neuromorphic systems are capable of categorization and decision making. Two existing applications utilize temporal logic, including LIDAR sensitivity that

is analogous to an owl's echolocation system and the escape response of a crayfish [18, 25] as detailed in Sect. 8.6. Combined with learning algorithms such as STDP which has recently been demonstrated in optics [68], networks could potentially perform more complex tasks such as spike-pattern cluster analysis.

8.10 Summary and Concluding Remarks

In an effort to break the limitations inherent in traditional von Neumann architectures, some recent projects in computing have sought more effective signal processing techniques by leveraging the underlying physics of devices [3, 4, 6, 69–71]. Cognitive computing platforms inspired by biological neural networks could solve unconventional computing problems and outperform current technology in both power efficiency and complexity [72–74]. These novel systems rely on alternative sets of computational principles, including hybrid analog-digital signal representations, co-location of memory and processing, unsupervised learning, and distributed representations of information.

On the cellular level, the brain encodes information as events or *spikes* in time [10], a hybrid signal with both analog and digital properties. Studies of morphology and physiology have pinpointed the LIF model as an effective spiking model to describe a variety of different biologically observed phenomena [2]. From the standpoint of computability and complexity theory, LIF neurons are powerful computational primitives that are capable of simulating both Turing machines and traditional sigmoidal neural networks [15]. These units perform a small set of basic operations (delaying, weighting, spatial summation, temporal integration, and thresholding) that are integrated into a single device capable of implementing a variety of processing tasks, including binary classification, adaptive feedback, and temporal logic. Though purely analog processing is susceptible to noise accumulation, this is overcome by neurons because the amplitude of spikes carry no information. The digital spike can be thresheld and restored by each neuron, which can then be cascaded like electronic logic gates. Thus neurons exploit both the bandwidth efficiency of analog signal processing and the scalability of digital logic. Networks of connected neurons can perform a wide variety of very powerful computing systems.

Spike processing algorithms are well understood in a number of important biological sensory processing systems and are finding growing use in signal processing applications [15]. The combination of these physiological principles with engineering not only helps in studying biological neural circuits [75], but also opens up a wide range of applications in emulating biological circuits *in silico*. A photonic realization of spiking neuron dynamics harnesses the high-switching speeds, wide communication bandwidth of optics, and low cross-talk achievable in photonics, making them well suited for an ultrafast spike-based information scheme. Optical spike processing devices are a result of the cross-fertilization between spike processing in the fields of computational neuroscience and high speed processing in the nonlinear optical device physics. While the fastest timescale on which biological neurons operate is in

the order of milliseconds, the photonic integrate-and-fire devices operate on picosecond width pulses, and have an integration time constant in the order of 100 ps, which is about eight orders of magnitude faster. Because of this, photonic spike processors can access a computational domain that is inaccessible by other technologies. This domain, which we describe as *ultrafast cognitive computing*, represents an unexplored processing paradigm that could have a wide range of applications in adaptive control, learning, perception, motion control, sensory processing, autonomous robotics, and cognitive processing of the radio frequency spectrum.

Recently, there has been a growing interest in photonic spike processing which has spawned a rich search for an appropriate computational primitive. In this context, we have demonstrated one of the first implementations of photonic spiking neurons capable of processing signals at terahertz rates [16, 17, 39]. Like its physiological counterpart, functions performed by photonic neurons are determined by the configuration of a set of parameters including the weights and delays of interconnections, the temporal integration time, and the spiking threshold. The first implementations of a photonic spiking neuron [16] achieves noise suppression and thresholding through a NOLM, generated pulses synchronously via a mode-locked laser, and utilized a SOA for integration. This model demonstrated temporal integration and spike thresholding, but too much excitation could lead to the release of multiple pulses, degrading spike-encoded information. In addition, spikes were not asynchronous, making the output of the system digital. Next, we demonstrated a fully functioning photonic neuron [17] which integrated both excitation and inhibition, but also suffered from the problems mentioned above. A newer asynchronous model was proposed recently by our group in [39] that generated spikes based on incoming spikes. Although the system could emit a pulse at analog times, because the system did not generate its own spikes through internal mechanisms, the spikes could eventually degrade into noise as non-spike inputs led to non-spike outputs.

These implementations are based on discrete photonic components and together with the reliance on nonlinear fibers and other similar technologies, have made these demonstrations bulky (on the order of meters), complex, and power-hungry (hundreds of watts). These platforms are simply unscalable beyond a few neurons and are best-suited for broadband signal processing applications that can be implemented with simple circuits. For example, we recently demonstrated several prototypical lightwave neuromorphic circuits including the simple auditory localization inspired by the barn owl [25], useful for light detection and ranging (LIDAR) localization, and the crayfish tail-flip escape response [18], which demonstrates accurate, picosecond pattern classification.

Integrated lasers, in contrast to first bench-top systems, are physically compact and are capable of using feedback rather than feedforward dynamics to radically enhance nonlinearity. Feedback allows for the emergence of more complex behaviors, including bistability, the formation of attractors, and excitability. The excitable models are a step in the right direction, given the similarities between lasers and biological phenomena and the stability of feedback systems.

Our latest photonic neuron model—excitable LIF laser neuron—avoids many previous issues by combining an excitable approach with some of the ideas of the

feedforward model. The integration of a laser gain section—which is dynamically analogous to an SOA—and a saturable absorber, which is essentially a feedback version of the thresholder used in the fiber model, leads to many desirable properties and a close analogy with biology, including the formation of highly stereotyped and well-defined optical spikes. Furthermore, this device is capable of implementing cortical-inspired algorithms. We have shown that, unlike previous models, our device can effectively perform cognitive algorithms at ultrafast time scales. Networks of such devices are stable, robust to noise, and can recognize patterns.

One of the greatest strengths of neural systems is their natural amenability to learning. Learning based on experience is essential for a system to dynamically adapt to an unpredictable or changing environment, or one that is too complex to be characterized *a priori*. There are several different methods by which neurons can learn: unsupervised learning extract structure from an unknown signal environment for efficient encoding, dimensionality reduction, or clustering; supervised learning programs a system to perform a function defined in terms of its input-output pairs; and reinforcement learning optimizes a particular function based on performance criterion. In all these cases, the learning mechanism configures a set of system parameters, such as synaptic weights, based upon the activity of the system in response to its inputs, and as a consequence, the system changes. Synaptic plasticity is not only responsible for learning and information storage but also the formation and refinement of neuronal circuits during development. With photonic neurons, we focus on learning through a type of synaptic plasticity called STDP, whereby the strength on input connections is determined by the relative timing of pre-synaptic and post-synaptic spikes. The experimentally demonstrated optical STDP [68] is potentially useful for applications including coincidence detection, sequence learning, path learning, and directional selectivity in visual response. These demonstration forms the foundation for photonic neurons to learn about the environment like a biological brain.

All the components employed for optical STDP—including the Mach-Zehnder configuration, the EAM, and the SOA—can be fabricated with a small footprint in planar photonics. Integrating this together with LIF excitable neurons on a single chip could lead to systems that emulate a well-established paradigm for adaptive computing on a scalable platform. These compact, adaptive, unconventional processing systems would operate on unprecedented time scales. Large networks could potentially be constructed as LSMs for optical reservoir computing [76] to aid in the study of biology or open up new applications to high bandwidth signal processing, ultrafast control loops, ultralow latency classification functions, and ultrafast adaptive processing.

Acknowledgments This work was supported by Lockheed Martin Advanced Technology Laboratory through the IRAD program, as well as the Lockheed Martin Corporation through the Corporate University Research Program. The authors also acknowledge the support of the NSF MIRTHE Center at Princeton University, the Pyne Fund and Essig Enright Fund for Engineering in Neuroscience. The work of M. A. Nahmias and A. N. Tait was supported by the National Science Foundation Graduate Research Fellowship (NSF-GRF). The work of B. J. Shastri was supported by the National Sciences and Engineering Research Council of Canada (NSERC) Postdoctoral Fellowship (PDF).

References

1. R. Sarpeshkar, Neural Comput. **10**(7), 1601 (1998)
2. C. Koch, Biophysics of Computation: Information Processing in Single Neurons (Computational Neuroscience) (Oxford University Press, 1998)
3. P. Merolla, J. Arthur, F. Akopyan, N. Imam, R. Manohar, D. Modha, in Custom Integrated Circuits Conference (CICC). IEEE **2011**, 1–4 (2011)
4. J. Seo, B. Brezzo, Y. Liu, B. Parker, S. Esser, R. Montoye, B. Rajendran, J. Tierno, L. Chang, D. Modha, et al., in Custom Integrated Circuits Conference (CICC), (IEEE, 2011), pp. 1–4
5. K. Likharev, A. Mayr, I. Muckra, Ö. Türel, Ann. N. Y. Acad. Sci. **1006**(1), 146 (2003)
6. G. Snider, Nanotechnology **18**(36), 365202 (2007)
7. Y. Abu-Mostafa, D. Psaltis, Sci. Am. **256**(3), 88 (1987)
8. S. Jutamulia, F. Yu, Opt. Laser Technol. **28**(2), 59 (1996)
9. M. Hill, E. Frietman, H. de Waardt, G. Khoe, H. Dorren, IEEE Trans. Neural Netw. **13**(6), 1504 (2002)
10. S. Thorpe, A. Delorme, R. Van Rullen et al., Neural Netw. **14**(6–7), 715 (2001)
11. W. Maass, Neural Netw. **10**(9), 1659 (1997)
12. D. Tal, E. Schwartz, Neural Comput. **9**(2), 305 (1997)
13. B. Lindner, L. Schimansky-Geier, A. Longtin, Phys. Rev. E **66**(3), 031916 (2002)
14. Y. Sakai, S. Funahashi, S. Shinomoto et al., Neural Netw. Official J. Int. Neural Netw. Soc. **12**(7–8), 1181 (1999)
15. W. Maass, C.M. Bishop (eds.), Pulsed neural networks (MIT Press, Cambridge, MA, 1999)
16. D. Rosenbluth, K. Kravtsov, M.P. Fok, P.R. Prucnal, Opt. Exp. **17**(25), 22767 (2009)
17. K. Kravtsov, M.P. Fok, D. Rosenbluth, P.R. Prucnal, Opt. Exp. **19**(3), 2133 (2011)
18. M.P. Fok, H. Deming, M. Nahmias, N. Rafidi, D. Rosenbluth, A. Tait, Y. Tian, P.R. Prucnal, Opt. Lett. **36**(1), 19 (2011)
19. Y. Tian, M. Fok, P. Prucnal, in 2011 Conference on IEEE Lasers and Electro-Optics (CLEO), (2011), pp. 1–2
20. K. Kravtsov, P.R. Prucnal, M.M. Bubnov, Opt. Express **15**(20), 13114 (2007)
21. A. Tait, M. Nahmias, M. Fok, P. Prucnal, in 2012 International Conference on Optical MEMS and Nanophotonics (OMN), (2012), pp. 212–213. doi:10.1109/OMEMS.2012.6318878
22. A.N.Tait et al., The DREAM: an integrated photonic thresholder
23. A.N. Tait, B.J. Shastri, M.P. Fok, M.A. Nahmias, P.R. Prucnal, The DREAM: An integrated photonic thresholder (accepted, 2013)
24. C. Koch, H. Li, *Vision Chips: Implementing Vision Algorithms With Analog Vlsi Circuits* (IEEE Press, New York, 1995)
25. P.R. Prucnal, M.P. Fok, D. Rosenbluth, K. Kravtsov, in ICO International Conference on Information Photonics (IP), (2011)
26. D. Young, *Nerve Cells and Animal Behaviour* (Cambridge University Press, Cambridge, 1989)
27. N. Edagawa, M. Suzuki, S. Yamamoto, IEICE Trans. Electron. E81-C(8), 1251 (1998)
28. S. Mahon, G. Casassus, C. Mulle, S. Charpier, J. Physiol **550**(Pt 3), 947 (2003)
29. D. Stellwagen, R.C. Malenka, Nature **440**(7087), 1054 (2006)
30. L.F. Abbott, S.B. Nelson, Nature Neurosci. Suppl. **3**, 1178 (2000)
31. C. Savin, P. Joshi, J. Triesch, PLoS Comput Biol **6**(4), e1000757 (2010)
32. B.J. Shastri, M.D. Levine, Mach. Vision Appl. **18**(2), 107 (2007)
33. J. Triesch, *in Proceedings of the 15th international conference on Artificial Neural Networks: biological Inspirations — Volume Part I, ICANN'05* (Springer, Berlin, 2005), pp. 65–70
34. G. Chechik, Neural Comput. **15**(7), 1481 (2003)
35. G. Indiveri, E. Chicca, R. Douglas, Trans. Neur. Netw. **17**(1), 211 (2006)
36. G.S. Snider, *in Proceedings of the 2008 IEEE International Symposium on Nanoscale Architecturess NANOARCH '08 (IEEE Computer Society* (Washington, DC, 2008), pp. 85–92
37. S.H. Jo, T. Chang, I. Ebong, B.B. Bhadviya, P. Mazumder, W. Lu, Nano Letters **10**(4), 1297 (2010)

38. G.S. Snider, SciDAC Rev. **10**, 58 (2008)
39. M.P. Fok, Y. Tian, D. Rosenbluth, P.R. Prucnal, Opt. Lett. **37**(16), 3309 (2012)
40. J.D. Joannopoulos, P.R. Villeneuve, S. Fan, Solid State Commun. **102**(2–3), 165 (1997)
41. E. Ozbay, Science **311**(5758), 189 (2006)
42. G.M. Wojcik, W.A. Kaminski, Neurocomputing **58–60**, 245 (2004)
43. M.A. Nahmias et al., A leaky integrate-and-fire laser neuron for ultrafast cognitive computing
44. M.A. Nahmias et al., An evanescent hybrid silicon laser neuron
45. B.J. Shastri et al. Exploring excitability in graphene for spike processing networks
46. B.J. Shastri et al., Graphene excitable laser for photonic spike processing
47. G. Spühler, R. Paschotta, R. Fluck, B. Braun, M. Moser, G. Zhang, E. Gini, U. Keller, JOSA B **16**(3), 376 (1999)
48. H. Wenzel, U. Bandelow, H. Wunsche, J. Rehberg, IEEE J. Quantum Electron **32**(1), 69 (1996)
49. D. Nugent, R. Plumb, M. Fisher, D. Davies, Electron. Lett. **31**(1), 43 (1995)
50. J. Dubbeldam, B. Krauskopf, Opt. commun. **159**(4), 325 (1999)
51. F. Koyama, J. Lightwave Technol. **24**(12), 4502 (2006)
52. S. Barbay, R. Kuszelewicz, A.M. Yacomotti, Opt. Lett. **36**(23), 4476 (2011)
53. Y. Li, T. Wang, R. Linke, Appl. Opt. **35**(8), 1282 (1996)
54. D. Taillaert, W. Bogaerts, P. Bienstman, T. Krauss, P. Van Daele, I. Moerman, S. Verstuyft, K. De Mesel, R. Baets, IEEE J. Quantum Electron **38**(7), 949 (2002)
55. D. Louderback, G. Pickrell, H. Lin, M. Fish, J. Hindi, P. Guilfoyle, Electron. Lett. **40**(17), 1064 (2004)
56. L. Coldren, S. Corzine, M. Mashanovitch, *Diode Lasers and Photonic Integrated Circuits* (New york, Wiley Series in Microwave and Optical Engineering (Wiley, 2011)
57. B.J. Shastri, C. Chen, K.D. Choquette, D.V. Plant, IEEE J. Quantum Electron **47**(12), 1537 (2011)
58. G.E. Giudice, D.V. Kuksenkov, H. Temkin, K.L. Lear, Appl. Phys. Lett. **74**(7), 899 (1999)
59. A. Hurtado, K. Schires, I. Henning, M. Adams, Appl. Phys. Lett. **100**(10), 103703 (2012)
60. A. Hurtado, I.D. Henning, M.J. Adams, Opt. Express 18(24), 25170 (2010) doi:10.1364/OE.18.025170. http://www.opticsexpress.org/abstract.cfm?URI=oe-18-24-25170
61. E. Izhikevich, IEEE Trans. Neural Netw. **15**(5), 1063 (2004)
62. L. Gelens, L. Mashal, S. Beri, W. Coomans, G. Van der Sande, J. Danckaert, G. Verschaffelt, arXiv, preprint arXiv:1108.3704 (2011)
63. W. Coomans, L. Gelens, S. Beri, J. Danckaert, G. Van der Sande, Phys. Rev. E **84**(3), 036209 (2011)
64. M. Herrmann, J.A. Hertz, A. Prugel-Bennett, Netw. Comput. Neural Syst. **6**(3), 403 (1995)
65. E.M. Izhikevich, Neural Comput. **18**(2), 245 (2006)
66. M. Abeles, *Corticonics: Neural Circuits of the Cerebral Cortex* (Cambridge University Press, Cambridge, 1991)
67. E. Bienenstock, Netw. Comput. Neural Syst. **6**(2), 179 (1995)
68. M.P. Fok, Y. Tian, D. Rosenbluth, P.R. Prucnal, Opt. Lett. (2012)
69. K. Boahen, Neuromorphic, Systems Engineering pp. 229–259 (1998)
70. J. Han, P. Jonker, Nanotechnology **14**(2), 224 (2003)
71. N. Mathur, Nature **419**(6907), 573 (2002)
72. G. Tononi, Biol. Bull. **215**(3), 216 (2008)
73. R. Ananthanarayanan, S.K. Esser, H.D. Simon, D.S. Modha, in Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis (ACM, New York, NY, USA, 2009), SC '09, pp. 63:1–63:12
74. D.S. Modha, R. Ananthanarayanan, S.K. Esser, A. Ndirango, A.J. Sherbondy, R. Singh, Commun. ACM **54**(8), 62 (2011)
75. L.S. Smith, A. Hamilton, Neuromorphic Systems: Engineering Silicon from Neurobiology, vol. 10 (World Scientific Publishing Company Incorporated, 1998)
76. K. Vandoorne, W. Dierckx, B. Schrauwen, D. Verstraeten, R. Baets, P. Bienstman, J.V. Campenhout, Opt. Exp. **16**(15), 11182 (2008)

Chapter 9
A Nanophotonic Computing Paradigm: Problem-Solving and Decision-Making Systems Using Spatiotemporal Photoexcitation Transfer Dynamics

Masashi Aono, Song-Ju Kim, Makoto Naruse, Masamitsu Wakabayashi, Hirokazu Hori, Motoichi Ohtsu and Masahiko Hara

Abstract In contrast to conventional digital computers that operate as instructed by programmers, biological organisms solve problems and make decisions through intrinsic spatiotemporal dynamics in which their dynamic components process

M. Aono (✉)
Earth-Life Science Institute, Tokyo Institute of Technology, 201202 Ookayama, Meguro-ku, Tokyo 152-8550, Japan
e-mail: masashi.aono@elsi.jp

S.- J. Kim (✉)
Atomic Electronics Group, WPI Center for Materials Nanoarchitectonics (MANA), National Institute for Materials Science (NIMS), 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan
e-mail: KIM.Songju@nims.go.jp

M. Naruse
Photonic Network Research Institute, National Institute of Information and Communications Technology, 4-2-1 Nukui-kita, Tokyo 184-8795, Koganei, Japan

M. Naruse · M. Ohtsu
Nanophotonics Research Center, Graduate School of Engineering, The University of Tokyo, 2-11-16 Yayoi, Bunkyo-ku 113-8656, Tokyo, Japan

M. Wakabayashi
Department of Biomolecular Engineering, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku 226-8501, Yokohama, Japan

H. Hori
Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, Takeda 4-3-11, Kofu 400-8511, Yamanashi, Japan

M. Ohtsu
Department of Electrical Engineering and Information Systems Graduate School of Engineering, The University of Tokyo, 2-11-16 Yayoi, Bunkyo-ku 113-8656, Tokyo, Japan

M. Hara
Department of Electronic Chemistry, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama 226-8503, Japan
e-mail: masahara@echem.titech.ac.jp

M. Naruse (ed.), *Nanophotonic Information Physics*,
Nano-Optics and Nanophotonics, DOI: 10.1007/978-3-642-40224-1_9,

environmental information in a self-organized manner. Previously, we formulated two mathematical models of spatiotemporal dynamics by which the single-celled amoeba (a plasmodial slime mold), which exhibits complex spatiotemporal oscillatory dynamics and sophisticated computing capabilities, could solve a problem and make a decision by changing its amorphous shape in dynamic and uncertain environments. These models can also be implemented by various physical systems that exhibit suitable spatiotemporal dynamics resembling the amoeba's shape-changing capability. Here we demonstrate that the photoexcitation transfer phenomena in certain quantum nanostructures mediated by optical near-field interactions mimic the amoeba-like spatiotemporal dynamics and can be used to solve two highly complex problems; the satisfiability problem, which is one of the most difficult combinatorial optimization problems, to determine whether a given logical proposition is self-consistent, and the multi-armed bandit problem, which is a decision-making problem in finding the most profitable option from among a number of options that provide rewards with different unknown probabilities. Our problem-solving and decision-making models exhibited better performances than conventionally known best algorithms. These demonstrations pave the way for a novel nanophotonic computing paradigm in which both coherent and dissipative processes are exploited for performing powerful solution searching and efficient decision-making with low energy consumption.

9.1 Introduction

"Natural computing" is an emerging research field that uses the knowledge obtained from various natural phenomena, including biological processes, to complement and overcome the limitations of conventional digital computers in solving intractable problems, making optimal decisions in uncertain environments, reducing energy consumption, and so on [1–4]. In this context, the single-celled amoeboid organism, a plasmodium of the true slime mold *Physarum polycephalum*, has been actively investigated owing to its intriguing computational capabilities. Nakagaki et al. showed that this amoeba, despite the absence of a central nervous system, connects the optimal routes among food sources by changing its amorphous shape [5, 6]. Aono et al. devised an amoeba-based computer (ABC) [7–9], which incorporated the amoeba to solve various optimization problems. These computational capabilities emerged from the amoeba's complex spatiotemporal behavior, in which the volume of each part oscillates in a fluctuating manner [10, 11].

ABC harnessed the complex spatiotemporal oscillatory dynamics of several pseudopod-like branches of the amoeba by introducing a unique optical feedback control, which was called "bounceback control." The organism inherently grows its branches by supplying its intracellular resource (protoplasm) to maximize its body area and thereby maximize its nutrient absorption from an agar plate. However, the branches retreat when stimulated by visible light as the resource bounces back from the illuminated region owing to the photoavoidance response. In ABC, we updated

the light stimulation conditions at regular intervals on the basis of a set of bounceback rules, which blocked the resource supply to undesirable combinations of the branches, with reference to the changes in the amoeba's shape. Under this bounceback control, the organism tried to deform into an optimal shape, maximizing the body area for maximal nutrient absorption while minimizing the risk of being exposed to light stimuli. Interestingly, ABC found a high-quality solution to the traveling salesman problem, which is one of the most studied combinatorial optimization problems, with high probability and in linearly suppressed time [12].

While the amoeba can change its shape, the conservation law of the total resource volume holds, which in effect enables the branches to rapidly exchange information on stimulated experiences in a spatially correlated manner. That is, a volume decrement in one branch is immediately compensated by volume increment(s) in the other branch(es). Extracting the essential factors by which the amoeba exhibited this powerful search ability, Aono et al. and Kim et al. formulated amoeba-inspired computing models for solving the satisfiability problem (SAT) [13] and the multi-armed bandit problem (BP) [14–17], respectively. These models suggested that to develop novel computing devices that operate much faster than the amoeba, it would be possible to use the stimulus-responsive spatiotemporal dynamics of various physical systems in which some resource of the system is transferred to its subsystems in a fluctuating manner. In fact, Naruse et al. showed that the spatiotemporal dynamics of photoexcitation transfer between quantum mechanical electronic states (excitonic states), which are implemented in semiconductor nanostructures and are mediated by optical near-field interactions, could be used to solve constraint satisfaction problems [18, 19], and they proposed a methodology to utilize photoexcitation transfer dynamics to solve SAT [20] and BP [21].

Optical near-field interactions occur at scales far below the wavelength of light and enable photoexcitation transfer to dipole-forbidden energy levels, which cannot be realized by conventional optical far-fields. A useful theoretical treatment of the near-field photoexcitation transfer process has been established on the basis of the dressed photon model [22], and the process has been experimentally demonstrated in quantum dot (QD) systems based on various semiconductors such as InGaAs [23], ZnO [24], and CdSe [25]. It has been demonstrated that photoexcitation transfer can be realized using two layers of two-dimensionally ordered InGaAs QDs at room temperature [26]. A system with a total QD density of 4.73×10^{12} /cm^2 has been fabricated using 60 highly stacked layers of InAs QDs [27]. DNA-based self-assembly technology can also be employed for fabricating controlled nanostructures [28]. It should be emphasized that the minimum energy dissipation in photoexcitation transfer has been shown to be 10^4 lesser than that required for a bit flip in a CMOS logic gate in conventional electrically wired devices [29]. These facts suggest that by exploiting photoexcitation transfer dynamics, a novel computing paradigm can be implemented in highly integrated low-energy-use quantum nanostructures.

In this paper, we first introduce the concept of photoexcitation transfer between QDs. Then we review our developed models for solving SAT and BP.

9.2 Photoexcitation Transfer

As shown in Fig. 9.1, we consider two spherical QDs whose radii are r_S and r_L ($>r_S$), called as a small QD (QD_S) and a large QD (QD_L), respectively. Under irradiation by input light, an exciton (electron–hole pair) is generated in QD_S. Here the photoexcitation transfer phenomena between QD_S and QD_L are considered, i.e., the transitions of an exciton to the states specified by (q_1, q_2), where q_1 and q_2 are the orbital angular momentum quantum number and magnetic quantum number, respectively. The energy eigenvalues of the states are given as follows:

$$E_{(q_1,q_2)} = E_g + E_{ex} + \frac{\hbar^2 \alpha_{(q_1,q_2)}{}^2}{2mr^2} \quad (q_1 = 1, 2, 3, \ldots), \tag{9.1}$$

where E_g is the bandgap energy of the bulk semiconductor, E_{ex} is the exciton binding energy in the bulk system, m is the effective mass of the exciton, and $\alpha_{(q_1,q_2)}$ are determined from the boundary conditions, for example, $\alpha_{(q_1,0)} = q_1\pi$, $\alpha_{(1,1)} = 4.49$.

There is a resonance between the level with quantum number $(1, 0)$ in QD_S, denoted by S in Fig. 9.1a, and that with quantum number $(1, 1)$ in QD_L, denoted by L^{Upper}, if $r_L/r_S = 4.49/\pi \approx 1.43$. These energy levels S and L^{Upper} are in resonance and are connected by an interdot optical near-field interaction, $U_{S,L}$, which is given by a Yukawa-type potential

$$U_{S,L} = \frac{\nu \exp(-\mu \mathrm{dst}(S, L))}{\mathrm{dst}(S, L)}, \tag{9.2}$$

where $\mathrm{dst}(S, L)$ denotes the distance between QD_S and QD_L, and ν and μ are constants [22, 30]. In typical light–matter interactions via optical far fields, transitions to states specified by $(q_1, q_2) = (1, 1)$ are not allowed, because this is a dipole-forbidden energy level. However, in optical near-field interactions, because of the

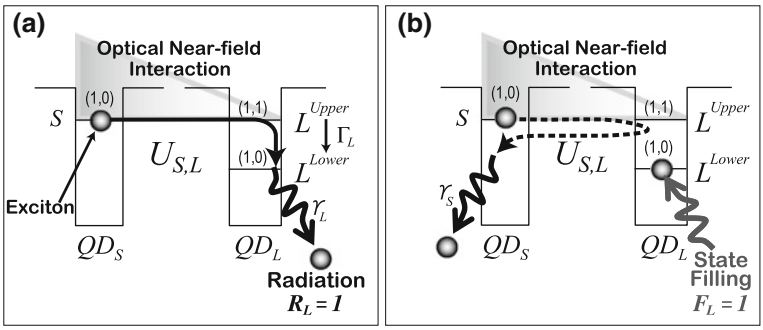


Fig. 9.1 Photoexcitation (excition) transfer between QD_S and QD_L [20]. **a** Exciton in QD_S is transferred to QD_L, from which it subsequently radiates. **b** The exciton is bounced back from QD_L when the lower energy level L^{Lower} is filled with another exciton

large spatial inhomogeneity of the localized optical near fields at the surface of nanoscale materials, L^{Upper} is allowed to be populated by excitons, violating the conventional optical selection rules [30]. Therefore, the exciton at level S in QD_S could be transferred to level L^{Upper} in QD_L.

In QD_L, because of the sublevel energy relaxation with the relaxation constant Γ_L, which is faster than the optical near-field interaction, the exciton relaxes to the $(1, 0)$ level, denoted by L^{Lower}, from where it radiatively dissipates (Fig. 9.1a). In addition, because the radiation lifetime of QDs is inversely proportional to their volume [31], we finally find "unidirectional" exciton transfer from QD_S to QD_L. We consider that the exciton is transferred from QD_S to QD_L when we observe light emission from QD_L due to the "radiation" of optical energy. The radiation from L^{Lower} is represented by the relaxation constant γ_L.

The unidirectionality of exciton transfer originates from the energy dissipation occurring in QD_L. Therefore, by disturbing the sublevel energy relaxation in QD_L, we can block exciton transfer to QD_L. In fact, when the lower energy level L^{Lower} of QD_L is filled with another exciton, the exciton in QD_S cannot move to QD_L. The blocked exciton will bounce back and forth between QD_L and QD_S (optical nutation) and will finally dissipate from QD_S according to the relaxation constant γ_S (Fig. 9.1b). We can fill the state L^{Lower} of QD_L by light stimulation; this is referred to as "state filling." The probability of exciton transfer to QD_L is reduced when it is state-filled, just like the branch of the amoeba that retreats when illuminated.

9.3 Nanophotonic Problem Solver

9.3.1 The Satisfiability Problem

The Satisfiability Problem (SAT) was the first problem proven to be nondeterministic polynomial time (NP)-complete, i.e., the most difficult problem among those that belong to the complexity class NP [32]. Given a logical formula ϕ, which consists of N boolean variables $x_i \in \{0\ (false), 1\ (true)\}$ $(i \in I = \{1, 2, \ldots, N\})$, SAT is the problem of determining whether there exists at least one "satisfying" assignment of the truth values (0 or 1) to the variables represented by x_i such that it makes the formula evaluate to true ($\phi = 1$). Roughly speaking, ϕ represents a logical proposition, and the existence of a satisfying assignment verifies that the proposition is self-consistent. For example, the formula $\phi_{ex} = (x_1 \vee \neg x_2) \wedge (\neg x_2 \vee x_3 \vee \neg x_4) \wedge (x_1 \vee x_3) \wedge (x_2 \vee \neg x_3) \wedge (x_3 \vee \neg x_4) \vee (\neg x_1 \vee x_4)$ has a satisfying assignment, $(x_1, x_2, x_3, x_4) = (1, 1, 1, 1)$, which is a uniquely existing solution.

Even if ϕ has more than one solution, this instance can be solved if at least one solution is found. However, to prove "unsatisfiability," the only sure method known to be applicable to arbitrary formulae is to check the inconsistency of all possible assignments, the number of which grows exponentially as 2^N. Therefore, all known SAT solvers are exponential-time algorithms.

SAT is called 3-SAT when ϕ consists of M clauses that are connected by $\wedge$ (and), and each clause connects at most three literals by $\vee$ (or) as $(x_j^* \vee x_k^* \vee x_l^*)$, where each literal x_i^* can be either x_i or $\neg x_i$. Any SAT instance can be transformed into a 3-SAT instance, and 3-SAT is also NP-complete. A powerful SAT solver has great potential for wide applications such as artificial intelligence, information security, and bioinformatics, because the NP-completeness implies that all NP problems, including thousands of practical real-world problems, can be reduced to SAT [32].

9.3.2 Spatiotemporal Dynamics of Photoexcitation Transfer

For simplicity, we consider a system where a QD_S is surrounded by four QD_L's, QD_A, QD_B, QD_C, and QD_D, each of which have the same upper level, lower level, sublevel relaxation constant, and radiation constant, L^{Upper}, L^{Lower}, Γ_L, and γ_L, as shown in Fig. 9.2. We suppose that the system initially has one exciton in S. For each QD_L, the exciton in S could be transferred to L^{Upper} through the interdot interaction $U_{S,L}$. Accordingly, we can derive quantum master equations using the density matrix formalism [30, 33]. The interaction Hamiltonian is given by

$$H_{int} = \begin{pmatrix} 0 & U_{S,A} & U_{S,B} & U_{S,C} & U_{S,D} \\ U_{S,A} & 0 & 0 & 0 & 0 \\ U_{S,B} & 0 & 0 & 0 & 0 \\ U_{S,C} & 0 & 0 & 0 & 0 \\ U_{S,D} & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{9.3}$$

Although interactions between the QD_Ls occur, for simplicity they are not considered here. The relaxation regarding the above-mentioned states is described by $N_\Gamma = diag(\gamma_S, \Gamma_A, \Gamma_B, \Gamma_C, \Gamma_D)$. Then, the Liouville equation for the system is

$$\frac{\mathrm{d}\rho(t)}{\mathrm{d}t} = -\frac{\mathrm{i}}{\hbar}[H_{int}, \rho(t)] - N_\Gamma \rho(t) - \rho(t) N_\Gamma, \tag{9.4}$$

where $\rho(t)$ is the density matrix with respect to the five energy levels and $\hbar$ is Planck's constant divided by 2π. Similarly, we can derive ordinary differential equations with respect to L^{Lower}, which is populated by the relaxations from L^{Upper} with constants Γ_L, which decay radiatively with relaxation constants γ_L. In the numerical calculation, we assume $U_{S,L}^{-1} = 100$ ps, $\Gamma_L^{-1} = 1$ ps, $\gamma_L^{-1} = 1$ ns, and $\gamma_S^{-1} = (r_L/r_S)^3 \times \gamma_L^{-1} \approx 2.92$ ns as a typical parameter set.

By numerically solving the above Liouville equations, the time evolution of "populations" of the lower energy levels of the QD_Ls, which are relevant to occurrences of radiation, can be calculated. We can obtain the probability p_L that the exciton in QD_S is transferred to QD_L, from which it subsequently radiates by numerically integrating the time evolution of the population of L_{Lower} over 6,000 ps and dividing it by a certain gain constant g, as shown in Fig. 9.2 a, b. In our numerical calculation,

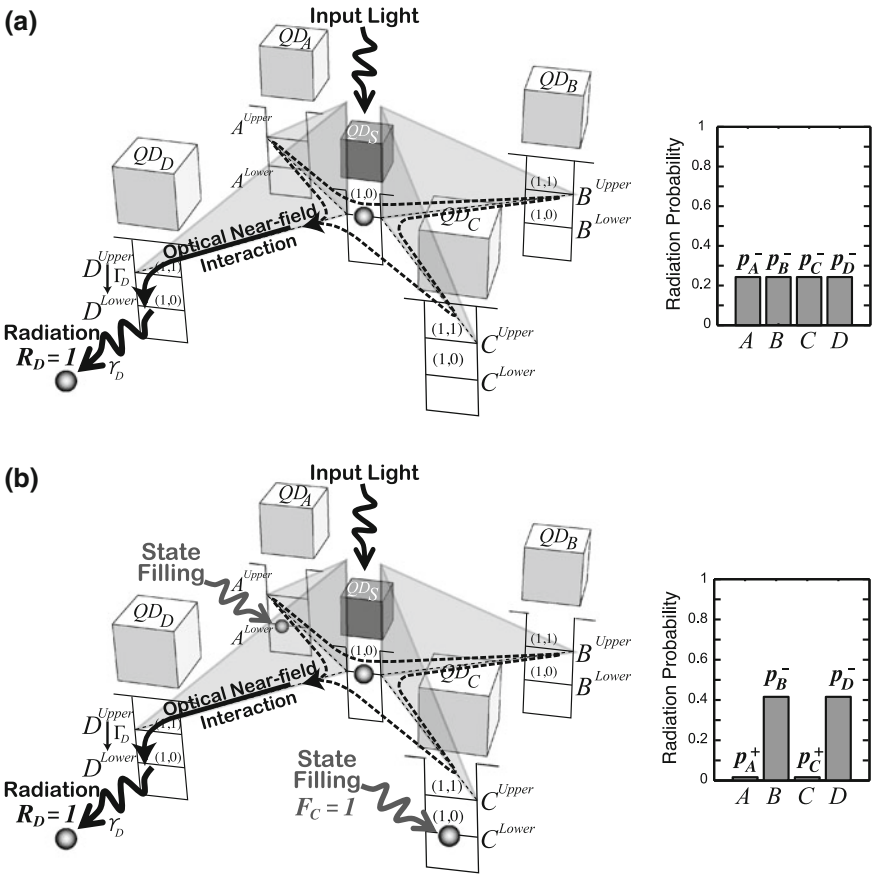


Fig. 9.2 Nanophotonic problem solver with a QD_S surrounded by four QD_Ls, QD_A, QD_B, QD_C, and QD_D [20]. **a** In the absence of state-filling stimulation, radiation occurs in the four large dots with an equal probability, as shown in the *right panel*. **b** When QD_A and QD_C are state-filled, the exciton is likely to be transferred to either QD_B or QD_D from which it radiates with a higher probability

we assume that radiation occurs in QD_L if a uniformly generated random number in [0.0, 1.0] is less than the value of p_L. Therefore, p_L represents the probability that radiation from QD_L is observed within 6,000 ps. Thus, more than one radiation event can occur in a number of QD_Ls. This verifies that $p_A + p_B + p_C + p_D \neq 1$. The radiation probabilities when QD_L is state-filled and nonstate-filled are denoted by p_L^+ and p_L^-, respectively.

The system shown in Fig. 9.2a uniformly grows the populations of A^{Lower}, B^{Lower}, C^{Lower}, and D^{Lower} while reducing the population of S and finally reaches an equilibrium. Figure 9.2b shows the case where QD_A and QD_C are subjected to state filling by light stimulation. A way of describing such a state-filling effect is to induce a

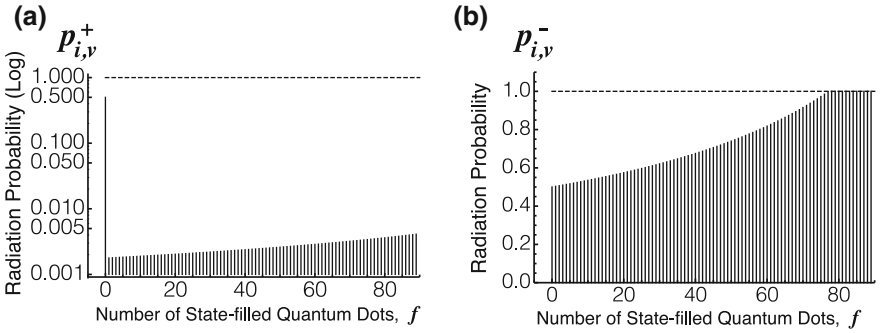


Fig. 9.3 Relationship between radiation probability and the number of all state-filled $QD_{j,u}$s, $f = \sum_{j,u} F_{j,u}$, in a nanophotonic problem solver consisting of 150 $QD_{j,u}$s for solving a 75-variable SAT [20]. **a** Radiation probability $p^{+}_{i,v}$ in a state-filled $QD_{i,v}$. **b** Radiation probability $p^{-}_{i,v}$ in a nonstate-filled $QD_{i,v}$. Each probability grows nonlinearly as a function of f, where $p^{-}_{i,v} = p^{+}_{i,v} = 0.5$ when $f = 0$

significant increase in the sublevel relaxation lifetime of the state-filled QD_A and QD_C; we assume that the lifetime increased to $\Gamma_A^{-1} = \Gamma_C^{-1} = 10^5$ ps. Such a formation has been validated in the literature [19]. Because of these changes in the parameters, the exciton is more likely to be transferred to QD_B or QD_D than to QD_A or QD_C, as shown in Fig. 9.2b.

As shown in the right-hand-side panels of Fig. 9.2 a, b, p_A and p_C decreased owing to state-filling stimulation, whereas p_B^- and p_D^- increased as if they tried to compensate for the decrements in p_A and p_C. That is, the radiation probability of each QD_L varied in response to the current state-filling stimulation applied to other distant QD_Ls. In other words, the stimulus response of each QD_L is not determined locally. This nonlocal property is shown more clearly in Fig. 9.3. The radiation probabilities of the state-filled and nonstate-filled QD_Ls, p_L^+ and p_L^-, increased nonlinearly as a function of the number of all state-filled QD_Ls. In the case of the amoeba, the conservation law of the total resource volume entailed a nonlocal correlation among the amoeba's branches; i.e., a volume increment in one branch is immediately compensated by volume decrement(s) in the other branch(es). This nonlocal correlation was shown to be useful for efficient and adaptive decision making [14].

9.3.3 Model of a Nanophotonic Problem Solver (NanoPS)

We call our numerical calculation model for simulating a nanophotonic problem solver "NanoPS." As shown in Fig. 9.4, to solve an N-variable 3-SAT, we use $2N$ QD_Ls, $QD_{i,v}$s, which receive optical energy from QD_S; the label (i, v) indicates that value $v \in \{0, 1\}$ is assigned to variable x_i ($i \in I = \{1, 2, \ldots, N\}$), i.e., $x_i = v$. When the exciton in QD_S is transferred to $QD_{i,v}$ and subsequently radiation is observed

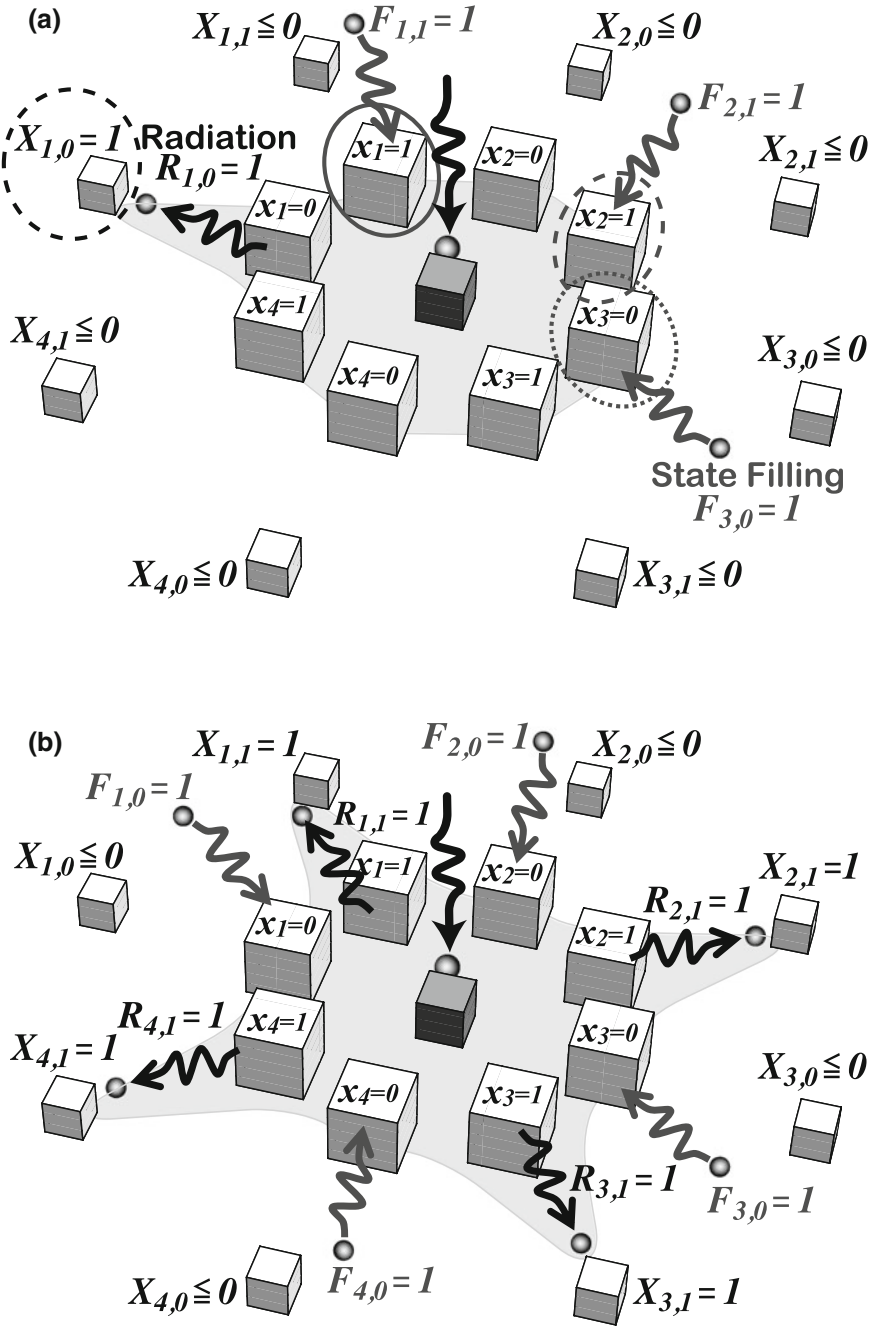


Fig. 9.4 Bounceback control and solution representation in a nanophotonic problem solver for solving the four-variable satisfiability problem [20]. **a** Bounceback control applies state-filling stimulations $F_{1,1}(t+1) = F_{2,1}(t+1) = F_{3,0}(t+1) = 1$, if $X_{1,0}(t) = 1$. **b** Configuration $X = (0, 1, 0, 1, 0, 1, 0, 1)$, which represents a solution $(x_1, x_2, x_3, x_4) = (1, 1, 1, 1)$

at a discrete time step t, we write this status as $R_{i,v}(t) = 1$, whereas $R_{i,v}(t) = 0$ indicates that no radiation occurs. When state-filling stimulation is applied to $QD_{i,v}$, we denote this status as $F_{i,v}(t) = 1$, whereas $F_{i,v}(t) = 0$ denotes no state-filling. Thus, radiation $R_{i,v}(t) = 1$ occurs with a probability that depends on the state-filling stimulation $F_{i,v}(t)$ as follows:

$$R_{i,v}(t) = \begin{cases} 1 \text{ (with a probability } p^{+}_{i,v}(t) \text{ if } F_{i,v}(t) = 1), \\ 1 \text{ (with a probability } p^{-}_{i,v}(t) \text{ if } F_{i,v}(t) = 0), \\ 0 \text{ (otherwise)}. \end{cases} \tag{9.5}$$

Figure 9.3 shows the dependence of the radiation probabilities of nonstate-filled and state-filled $QD_{i,v}$s, $p^{-}_{i,v}(t)$ and $p^{+}_{i,v}(t)$, on the number of all state-filled $QD_{j,u}$s, i.e., $\sum_{j,u} F_{j,u}(t)$.

Each radiation event $R_{i,v}$ is accumulated by a newly introduced variable $X_{i,v} \in \{-1, 0, 1\}$ as follows:

$$X_{i,v}(t+1) = \begin{cases} X_{i,v}(t) + 1 & (\text{if } R_{i,v}(t) = 1 \text{ and } X_{i,v}(t) < 1), \\ X_{i,v}(t) - 1 & (\text{if } R_{i,v}(t) = 0 \text{ and } X_{i,v}(t) > -1), \\ X_{i,v}(t) & (\text{otherwise}). \end{cases} \tag{9.6}$$

The above dynamics can be implemented either physically or digitally; i.e., the values of $X_{i,v}$ can be stored either by some additional QDs or by some external control unit, respectively. At each step t, the system transforms a configuration $X = (X_{1,0}, X_{1,1}, X_{2,0}, X_{2,1}, \ldots, X_{N,0}, X_{N,1})$ into an assignment $x = (x_1, x_2, \ldots, x_N)$ as follows:

$$x_i(t) = \begin{cases} 0 & (\text{if } X_{i,0}(t) = 1 \text{ and } X_{i,1}(t) \leq 0), \\ 1 & (\text{if } X_{i,0}(t) \leq 0 \text{ and } X_{i,1}(t) = 1), \\ x_i(t-1) & (\text{otherwise}), \end{cases} \tag{9.7}$$

where $x_i(0) = \textit{undefined}$ for all i.

The state-filling stimulations $F_{i,v}$s are updated synchronously according to the following dynamics:

$$F_{i,v}(t+1) = \begin{cases} 1 \left(\text{if } \exists (P, Q) \in B \left(\forall (j, u) \in P \, (X_{j,u}(t) = 1) \text{ and } (i, v) \in Q \right) \right), \\ 0 \text{ (otherwise)}, \end{cases} \tag{9.8}$$

where B is a set of bounceback rules to be defined in this section. Each element (P, Q) in B implies the following statement: if all the $X_{j,u}$s specified by P are positive at t, then stimulate all $QD_{i,v}$s specified by Q to inhibit their radiation at $t + 1$. Stated simply, if $x_j = u$, then x_i should not be v.

To understand the meaning of the bounceback rules, let us consider the example formula ϕ_{ex}. To satisfy this formula for $\phi_{ex} = 1$, we should make every clause *true* because all clauses are connected by $\wedge$. For example, suppose the system tried to

assign $x_1 = 0$, i.e., $X_{1,0}(t) = 1$, as indicated by the black broken circle in Fig. 9.4a. Now let us focus on the first clause $(x_1 \vee \neg x_2)$ in ϕ_{ex}. To make this clause *true*, if $x_1 = 0$ then x_2 should not be 1. Therefore, we apply state-filling stimulation $F_{2,1}(t+1) = 1$ to inhibit radiation $R_{2,1}(t+1)$ of $QD_{2,1}$, as indicated by the gray broken circle. On the other hand, because x_3 in the third clause $(x_1 \vee x_3)$ should not be 0, we apply $F_{3,0}(t+1) = 1$ (the gray dotted circle). In addition, we apply $F_{1,1}(t+1) = 1$ (the solid circle), because if $x_1 = 0$, obviously x_1 should not be 1. Likewise, the set of all bounceback rules B is determined by scanning all clauses in ϕ_{ex}.

The elements of the set $B = INTRA \cup INTER \cup CONTRA$ are formally defined as follows. *INTRA* forbids each variable i to take two values 0 and 1 simultaneously:

$$INTRA = \{\ (\{(i, v)\}, \{(i, 1-v)\}) \mid i \in I \wedge v \in \{0, 1\}\ \}. \tag{9.9}$$

Each clause $c = (x_j^* \vee x_k^* \vee x_l^*)$ in ϕ is represented as a set $C = \{j^*, k^*, l^*\}$ with its literals x_i^* mapped to $i^* = i$ if $x_i^* = x_i$ and to $-i$ otherwise, and the formula ϕ is expressed equivalently by a set Φ, which includes all the clauses as its elements. For example, the example formula ϕ_{ex} is transformed into $\Phi_{ex} = \{\{1, -2\}, \{-2, 3, -4\}, \{1, 3\}, \{2, -3\}, \{3, -4\}, \{-1, 4\}\}$. For each C in Φ and each variable i in C, *INTER* blocks the radiation [either $R_{i,0}(t+1)$ or $R_{i,1}(t+1)$] that makes c *false*:

$$INTER = \{(P, \{(i, 0)\}) \mid i \in C \wedge C \in \Phi\ \} \cup \{(P, \{(i, 1)\}) \mid -i \in C \wedge C \in \Phi\}, \tag{9.10}$$

where $P = \{\ (j, 0) \mid j \in C \wedge j \neq i\ \} \cup \{\ (j, 1) \mid -j \in C \wedge j \neq i\ \}$. Some rules in *INTER* may imply that neither 0 nor 1 can be assigned to a variable. To avoid this contradiction, for each variable i, we build *CONTRA* by checking all the relevant rules in *INTER*:

$$\begin{aligned} CONTRA = \{(P \cup P', P \cup P') \mid\ & i \in I \wedge (P, \{(i, 0)\}) \\ & \in INTER \wedge (P', \{(i, 1)\}) \in INTER\}. \end{aligned} \tag{9.11}$$

Before the computation, B is obtained in a polynomial time of $O(N \cdot M)$ by generating all the bounceback rules in *INTRA*, *INTER*, and *CONTRA* according to the above procedures.

The computation starts at $X_{i,v}(0) = R_{i,v}(0) = F_{i,v}(0) = 0$ for all (i, v), and the time evolution of the system is simulated by calculating the above equations iteratively. Figure 9.5 shows that the system successfully found the solution of the example formula ϕ_{ex} at $t = 13$, which is represented by the configuration shown in Fig. 9.4b.

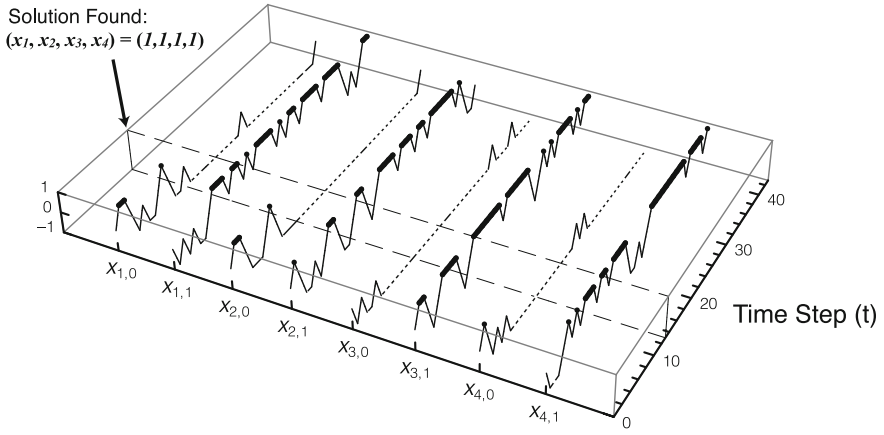


Fig. 9.5 Simulated time evolution of $X_{i,v}(t)$ in a nanophotonic problem solver [20]. *Black dots* and *broken dotted lines* indicate $X_{i,v}(t) = 1$ and $F_{i,v}(t) = 1$, respectively. The system found the solution $(x_1, x_2, x_3, x_4) = (1, 1, 1, 1)$ at $t = 13$

9.3.4 WalkSAT Algorithm

The performance of NanoPS is compared with that of the best studied stochastic local search algorithm, WalkSAT. Initially, an assignment $x(0) = (x_1(0), x_2(0), \ldots, x_N(0))$ is randomly chosen. At each time step t, by checking whether each clause in f is satisfied by the current assignment $x(t)$, WalkSAT randomly chooses one of the unsatisfied clauses and satisfies it by flipping one of its variables chosen at random as $x_i(t + 1) = 1 - x_i(t)$. This routine is iterated until a satisfying assignment is obtained, or we run out of time. Schöning estimated the average number of iterations that WalkSAT required for finding a solution to a 3-SAT problem as the exponential function of $(4/3)^N poly(N)$ [34]; WalkSAT is one of the fastest stochastic local search algorithms for solving this problem [35].

9.3.5 NanoPS versus WalkSAT

We used benchmark problem instances provided by SATLIB online [36], which were the most difficult 3-SAT instances obtained by randomly generating three-literal conjunctive normal form formulae, where the difficulty can be maximized by setting the ratio between the number of variables N and the number of clauses M at the phase transition region around $M/N = 4.26$ [37, 38]. We chose 100 instances from each of the test sets uf75–325 and uf100–430, which took satisfiable $N = 75$–$M = 325$ and $N = 100$–$M = 430$ formulae from the most difficult region, $M/N \approx 4.333$ and $M/N = 4.3$, respectively.

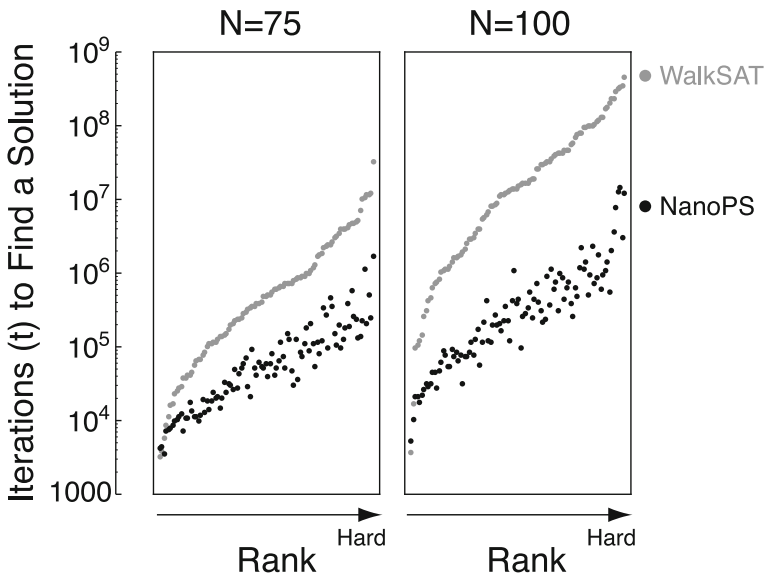


Fig. 9.6 Performance comparison between a simulation model of a nanophotonic problem solver, NanoPS (*black points*), and the well-known WalkSAT (*gray points*) for benchmark 3-SAT instances of $N = 75$ and $N = 100$ [20]. Each point indicates the average number of iterations over 500 trials required to find a solution for a particular instance. For each algorithm and each N, we evaluated 100 points (instances). The instances were sorted (ranked) from easiest to most difficult in the ascending order of the average number of iterations that WalkSAT required to find a solution. That is, the result of WalkSAT and that of NanoPS for an identical instance are located at the same position on the horizontal axis. The results are compared on a logarithmic scale, which implies that NanoPS has a significant advantage over WalkSAT

For each instance, we conducted 500 trials consisting of Monte Carlo simulations to obtain the average number of iterations (time steps t) required to find a solution. As shown in Fig. 9.6, NanoPS found a solution after a much smaller number of iterations than WalkSAT. The advantage of NanoPS over WalkSAT increased as N increased.

9.4 Nanophotonic Decision Maker

9.4.1 Multi-Armed Bandit Problem (BP)

For a formal discussion on the decision making, we tackle BP, which is stated as follows. Consider a number of slot machines, each of which rewards the player with a coin at a certain probability P_k ($k \in \{1, 2, \ldots, N\}$) when played. To maximize the total amount of reward, it is necessary to make a quick and accurate judgment of which machine has the highest probability of giving a reward. To accomplish

this, the player should gather information about many machines; however, in this process, the player should not fail to exploit the reward from the known best machine. These requirements are not easily met simultaneously, because there is a trade-off between "exploration" and "exploitation." BP is used to determine the optimal strategy for maximizing the total reward with incompatible demands, either by exploiting rewards obtained through already collected information or by exploring new information to acquire higher pay-offs through risk taking. Biological organisms commonly encounter this "exploration-exploitation dilemma" in their struggle to survive in an unknown world. This dilemma has no known generally optimal solution.

BP was originally described by Robbins [39], although the same problem in essence was also studied by Thompson [40]. However, the optimal strategy is known only for a limited class of problems in which the reward distributions are assumed to be known to the players [41, 42]. Furthermore, computing the Gittins index in practice is not tractable for many problems. Agrawal and Auer et al. proposed algorithms that could express the index as a simple function of the total reward obtained from a machine [43, 44]. These algorithms are used worldwide for many applications, such as Monte Carlo tree searches [45, 46].

Inspired by the amoeba's shape-changing process under dynamic light stimuli, Kim et al. proposed an algorithm for BP called the "tug-of-war model" (TOW) [14–17]. TOW is a dynamical system model of an amoeba-like body, which maintains a constant intracellular resource volume while collecting environmental information by concurrently expanding and shrinking its branches. The conservation law entails a "nonlocal correlation" among the branches. The volume increment in one branch is immediately compensated by volume decrement(s) in the other branch(es). This nonlocal correlation was shown to enhance the performance in solving BP [14].

9.4.2 Model of a Nanophotonic Decision Maker (NanoDM)

We describe a numerical calculation model for simulating a nanophotonic decision maker, called "NanoDM," which implements a variant of TOW using photoexcitation transfer dynamics among QDs. Although we only demonstrate only the two-armed case, NanoDM can be easily extended to N-armed ($N > 2$) cases. We use three types of cubic QDs with side lengths of a, $\sqrt{2}a$ and $2a$, which are represented by QD_S, QD_M and QD_L respectively, as shown in Fig. 9.7. We assume that five QDs are one-dimensionally arranged in QD_L-QD_M-QD_S-QD_M-QD_L or QD_M-QD_L-QD_S-QD_L-QD_M. When an exciton is generated at QD_S, it is transferred to the lowest energy levels in QD_Ls; we observe negligible radiation from QD_Ms. However, when the lowest energy levels of QD_Ls are populated with excitons by control lights, which induce state-filling effects, the exciton at QD_S is more likely to be radiated from QD_M.

We consider the radiation from either left QD_{ML} or right QD_{MR} as the decision of selecting slot machine A and B, respectively. The intensity of the control light to

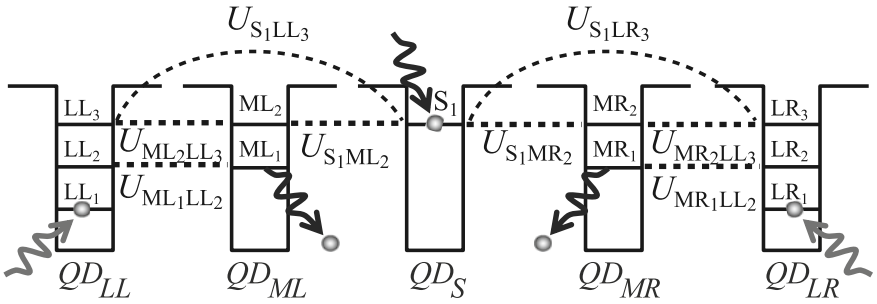


Fig. 9.7 Photoexcitation (exciton) transfer in a nanophotonic decision maker consisting of five QDs denoted as QD_{LL}, QD_{ML}, QD_S, QD_{MR} and QD_{LR} [21]. Two cubic quantum dots QD_S and QD_M, whose side lengths are a and $\sqrt{2}a$, respectively, are located close to each other. The exciton in QD_S can be transferred to neighboring structures QD_M via optical near-field interactions, denoted by $U_{S_1M_2}$, because there exists a resonance between the level of quantum number (1, 1, 1) for QD_S (denoted by S_1) and that of quantum number (2, 1, 1) for QD_M (M_2). The energy levels in the system are summarized as follows. The (2, 1, 1)-levels of QD_{ML}, QD_{MR}, QD_{LL} and QD_{LR} are denoted by ML_2, MR_2, LL_2 and LR_2, respectively. The (1, 1, 1)-level of QD_{ML}, QD_{MR}, QD_{LL} and QD_{LR} are denoted by ML_1, MR_1, LL_1 and LR_1, respectively. The (2, 2, 2)-levels of QD_{LL} and QD_{LR} are denoted by LL_3 and LR_3, respectively. The optical near-field interactions are $U_{S_1Mi_2}$, $U_{S_1Li_3}$, $U_{Mi_2Li_3}$ and $U_{Mi_1Li_2}$ ($i = L, R$)

induce state-filling at the left and right QD_Ls is respectively modulated on the basis of the resultant rewards obtained from the chosen slot machine. It should be noted that photoexcitation transfer phenomena among QDs are fundamentally probabilistic and are described on the basis of density matrix formalism in a manner that is similar to NanoPS. Until energy dissipation is induced, an exciton simultaneously interacts with potentially transferable destination QDs in the resonant energy level. We exploit such probabilistic behavior for the function of exploration for decision making.

NanoDM is modeled using density matrix formalism. For simplicity, we assume one excitation system. There are in total 11 energy levels in the system: S_1 in QD_S; ML_1 and ML_2 in QD_{ML}; LL_1, LL_2 and LL_3 in QD_{LL}; MR_1 and MR_2 in QD_{MR}; LR_1, LR_2 and LR_3 in QD_{LR}. Therefore the number of different states occupying these energy levels is 12 including the vacancy state, as schematically shown in Fig. 9.8. Because a fast intersublevel transition in QD_Ls and QD_Ms is assumed, it is useful to establish theoretical treatments on the basis of the exciton population in the system composed of QD_S, QD_Ms and QD_Ls, where 11 basis states are assumed.

The optical near-field interaction between energy levels E_1 and E_2 is represented by $U_{E_1E_2}$. For instance, the interaction between the (1, 1, 1)-level of QD_S (S_1) and the (2, 1, 1)-level of QD_{ML} (ML_2) is denoted by $U_{S_1ML_2}$. The radiative relaxation rates from S_1, Mi_1, and Li_1 are given by γ_{S_1}, γ_{Mi_1} and γ_{Li_1}, respectively. The quantum master equation of the total system is then given by [33]

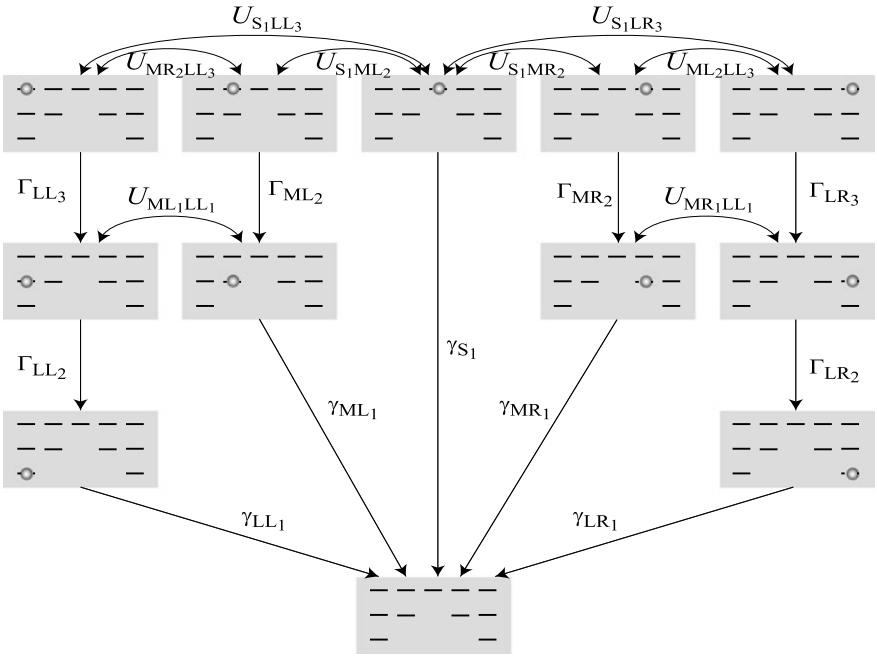


Fig. 9.8 Schematic summary of the state transitions in nanophotonic decision maker[21]. The relaxation rates Γ_{LL_3}, Γ_{LL_2}, Γ_{ML_2}, Γ_{MR_2}, Γ_{LR_3} and Γ_{LR_2}, and the radiative decay rates γ_{LL_1}, γ_{ML_1}, γ_{S_1}, γ_{MR_1} and γ_{LR_1} are shown

$$\begin{aligned}\frac{d\rho(t)}{dt} = &-\frac{\mathrm{i}}{\hbar}[H_{int}, \rho(t)] \\ &+ \sum_{i=S_1, ML_1, MR_1, LL_1, LR_1} \frac{\gamma_i}{2}\left(2R_i\rho(t)R_i^\dagger - R_i^\dagger R_i\rho(t) - \rho(t)R_i^\dagger R_i\right) \\ &+ \sum_{i=ML_2, LL_3, LL_2, MR_2, LR_3, LR_2} \frac{\Gamma_i}{2}\left(2S_i\rho(t)S_i^\dagger - S_i^\dagger S_i\rho(t) - \rho(t)S_i^\dagger S_i\right),\end{aligned} \tag{9.12}$$

where the interaction Hamiltonian is given by H_{int}. Let the (i, i) and (j, j) elements of $\rho(t)$ be the probabilities of the two states that are transformable between each other via an optical near-field interaction denoted by U_{ij}. Then, the (i, j) and (j, i) elements of the interaction Hamiltonian are given by U_{ij}. The matrices R_i ($i = S_1, ML_1, MR_1, LL_1, LR_1$) are annihilation operators that annihilate excitations in S_1, ML_1, MR_1, LL_1 and LR_1, respectively, via radiative relaxations. The matrices $R_i^\dagger$ ($i = S_1, ML_1, MR_1, LL_1, LR_1$) are creation operators given by the transposes of the matrices of R_i. The radiative decay times $\gamma_{LL_1}^{-1} = \gamma_{LR_1}^{-1} = 1$ ns, $\gamma_{ML_1}^{-1} = \gamma_{MR_1}^{-1} = 2^{\frac{3}{2}} \times 1 \sim 2.83$ ns and $\gamma_{S_1}^{-1} = 2^3 \times 1 \sim 8$ ns are inversely proportional to the volumes of the QD_S.

The matrices S_i ($i = ML_2, LL_3, LL_2, MR_2, LR_3, LR_2$) are annihilation operators that annihilate excitations in $ML_2, LL_3, LL_2, MR_2, LR_3$ and LR_2, respectively, via sublevel relaxations Γ_i and populate excitations in the corresponding lower energy levels.

When there is no state-filling induced at LL_1 and LR_1, the sublevel relaxations are assumed to be equally fast, $\Gamma_i = 10$ ps. On the other hand, when control lights are used to induce state-filling at LL_1 and/or LR_1, we have to consider a multi-excitation system in order to calculate the exact dynamics of the system. Another way of describing the effect of state filling is that the sublevel relaxation time increases, for example by a factor of 10, that is, $\Gamma_{LL_2}^{-1} = \Gamma_{LR_2}^{-1} = 10 \times 10$ ps. Such an approach has been validated in [19].

Through optical near-field interactions, an exciton generated at QD_S is transferred to the lowest energy levels in the largest-size QD, namely the energy level LL_1 or LR_1. However, when LL_1 and LR_1 are occupied by other excitations, the input excitation generated at S_1 should be relaxed from the middle-sized QD, that is, ML_1 and MR_1. The concept of the NanoDM is to induce state filling at LL_1 and/or LR_1 while observing radiations from ML_1 and MR_1. If radiation occurs in ML_1 (MR_1), then we consider that the system selects machine A (B). We can modulate these radiation probabilities by changing the intensity of the incident light.

9.4.3 Intensity Adjuster of NanoDM

We adopt the intensity adjuster (IA) to modulate the intensity of incident light to large QDs, as shown at the bottom of Fig. 9.9. The initial position of the IA is zero. In this case, the same intensity of light is applied to both LL_1 and LR_1. If we move the IA to the right, the intensity at the right increases and that at the left decreases. In contrast, if we move the IA to the left, the intensity at the left increases and that at the right decreases. This situation can be described by the following relaxation rate parameters as functions of the IA position j: $\Gamma_{LR_2} = \frac{1}{100} - \frac{1}{10{,}000}\, j + \frac{1}{100{,}000}$, and $\Gamma_{LL_2} = \frac{1}{100} + \frac{1}{10{,}000} j + \frac{1}{100{,}000}$.

We calculated two (right and left) radiation probabilities in advance from ML_1 and MR_1, namely $S_A(j)$ and $S_B(j)$, in each of the 201 states ($-100 \le j \le 100$) by using the quantum master equation of the entire system. We used the following relaxation rate and radiative decay parameters as shown in Fig. 9.8, $\gamma_{S_1} = \frac{1}{\left(2\right)^3 1{,}000}$, $\gamma_{MR_1} = \frac{1}{\left(\sqrt{2}\right)^3 1{,}000}$, $\gamma_{ML_1} = \frac{1}{\left(\sqrt{2}\right)^3 1{,}000}$, $\gamma_{LR_1} = \frac{1}{1{,}000}$, $\gamma_{LL_1} = \frac{1}{1{,}000}$, $\Gamma_{MR_2} = \frac{1}{10}$, $\Gamma_{LR_3} = \frac{1}{10}$, $\Gamma_{LR_2} = \frac{1}{100} - \frac{1}{10{,}000}\, j + \frac{1}{100{,}000}$, $\Gamma_{ML_2} = \frac{1}{10}$, $\Gamma_{LL_3} = \frac{1}{10}$ and $\Gamma_{LL_2} = \frac{1}{100} + \frac{1}{10{,}000}\, j + \frac{1}{100{,}000}$, where j represents the position of the IA.

Note that we used the following optical near-field interaction parameters so that the QDs are one-dimensionally arranged in the QD_M-QD_L-QD_S-QD_L-QD_M order for this calculation: $U_{S_1MR_2} = \frac{1}{10{,}000{,}000}$, $U_{S_1LR_3} = \frac{1}{100}$, $U_{MR_2LR_3} = \frac{1}{10{,}000}$, $U_{MR_1LR_2} = \frac{1}{10{,}000}$, $U_{S_1ML_2} = \frac{1}{10{,}000{,}000}$, $U_{S_1LL_3} = \frac{1}{100}$, $U_{ML_2LL_3} = \frac{1}{10{,}000}$ and $U_{ML_1LL_2} = \frac{1}{10{,}000}$. Here we used $a = 0.1$ and $k = 10^9$ for the parameters. Then, the distance

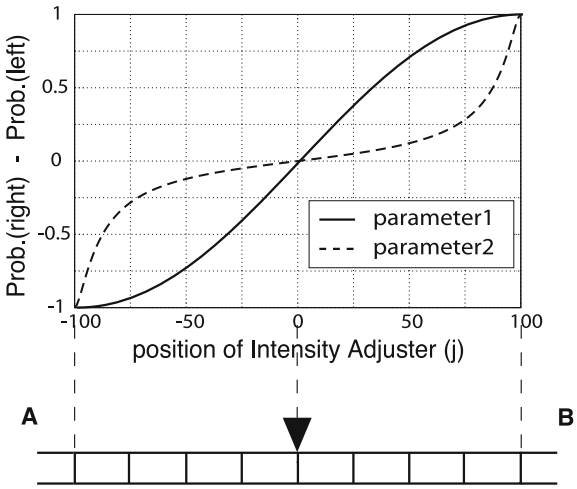


Fig. 9.9 Intensity adjuster (IA) and the difference between radiation probabilities from ML_1 and MR_1 in a nanophotonic decision maker [21]. The difference between radiation probabilities $S_B(j) - S_A(j)$ as a function of the IA position j, which are calculated from the quantum master equation of the total system, is denoted by the *solid line*. Here we used the parameters $\Gamma_{LR_2} = \frac{1}{100} - \frac{1}{10{,}000}\, j + \frac{1}{100{,}000}$, and $\Gamma_{LL_2} = \frac{1}{100} + \frac{1}{10{,}000}\, j + \frac{1}{100{,}000}$. As supporting information, the *dashed line* denotes the case where $\Gamma_{LR_2} = \frac{1}{10+500(100+j)}$, and $\Gamma_{LL_2} = \frac{1}{10-500(100-j)}$

between QD_S and QD_L (r_1) is 20 nm, and the distance between QD_L and QD_M (r_2) is 24.5 nm, such that $U(r_1) = \frac{1}{100}$ and $U(r_2) = \frac{1}{10{,}000}$. Finally, we obtained two radiation probabilities whose maximum ratio is 8520.82. The difference between radiation probabilities, $S_B(j) - S_A(j)$, is shown by the solid line in Fig. 9.9.

If $j > 0$, the intensity of incident light to the LR_1 increases (the Γ_{LR_2} decreases by the amount of $j/10{,}000$), while that to the LL_1 decreases (the Γ_{LL_2} increases by the amount of $j/10{,}000$). Correspondingly, the radiation probability $S_B(j)$ is larger than $S_A(j)$ for $j > 0$, while the radiation probability $S_A(j)$ is larger than $S_B(j)$ for $j < 0$. For $j < -100$, we used probabilities $S_A(-100)$ and $S_B(-100)$. Similarly, we used $S_A(100)$ and $S_B(100)$ for $j > 100$.

The dynamics of the IA are defined as follows:

1. Set the IA position j to 0.
2. Select machine A or B by using $S_A(j)$ and $S_B(j)$.
3. Play on the selected machine.
4. If a coin is dispensed, then move the IA to the selected machine's direction, that is, $j = j - D$ for A, and $j = j + D$ for B.
5. If no coin is dispensed, then move the IA to the inverse direction of the selected machine, that is, $j = j + D$ for A, and $j = j - D$ for B.
6. Repeat step (2).

Here, D is a parameter. In this way, NanoDM selects A or B, and the IA moves to the right or left according to the reward.

9.4.4 Softmax Algorithm

The performance of NanoDM is compared with that of the Softmax algorithm, which is known as the best-fitting algorithm for human decision-making behavior in the BP task [47]. In the Softmax algorithm, the probability of selecting A or B, $P'_A(t)$ or $P'_B(t)$, is given by the following Boltzmann distributions:

$$P'_A(t) = \frac{\exp[\beta Q'_A(t)]}{\exp[\beta Q'_A(t)] + \exp[\beta Q'_B(t)]}, \tag{9.13}$$

$$P'_B(t) = \frac{\exp[\beta Q'_B(t)]}{\exp[\beta Q'_A(t)] + \exp[\beta Q'_B(t)]}, \tag{9.14}$$

where Q'_k ($k \in \{A, B\}$) is the estimated reward probability of slot machines k, denoted by P_k, and "temperature" β is a parameter. These estimates are given as,

$$Q'_k(t) = \frac{W_k(t)}{M_k(t)}, \tag{9.15}$$

$$W_k(t) = w_k(t) + W_k(t-1), \tag{9.16}$$

$$M_k(t) = m_k(t) + M_k(t-1). \tag{9.17}$$

Here $w_k(t)$ ($k \in \{A, B\}$) is 1 if a reward is dispensed from machine k at time t, otherwise 0, and $m_k(t)$ ($k \in \{A, B\}$) is 1 if machine k is selected at time t, otherwise 0.

In our study, the "temperature" β was modified to a time-dependent form as follows:

$$\beta(t) = \tau \cdot t. \tag{9.18}$$

Here, τ is a parameter that determines the growth rate. $\beta = 0$ corresponds to a random selection, and $\beta \to \infty$ corresponds to a greedy action. A greedy action means that the player selects A if $Q'_A > Q'_B$, or selects B if $Q'_A < Q'_B$.

9.4.5 NanoDM versus Softmax

Figure 9.10a, b demonstrate the efficiency (cumulative rate of correct selections) for NanoDM (solid line) and Softmax with the optimized parameter τ (broken line) in the case where the reward probabilities of the slot machines are (a) $P_A = 0.2$ and $P_B = 0.8$ and (b) $P_A = 0.4$ and $P_B = 0.6$. In these cases, the correct selection is "B" because P_B is greater than P_A. These cumulative rates of correct selections are average values for each 1,000 samples. Hence, each value corresponds to the average number of coins acquired from the slot machines. Even with a nonoptimized parameter D, the performance of NanoDM was higher than that of Softmax with an optimized parameter τ, in a wide parameter range of $D = 10 - 100$ although we show only the $D = 50$ case in Fig. 9.10a, b.

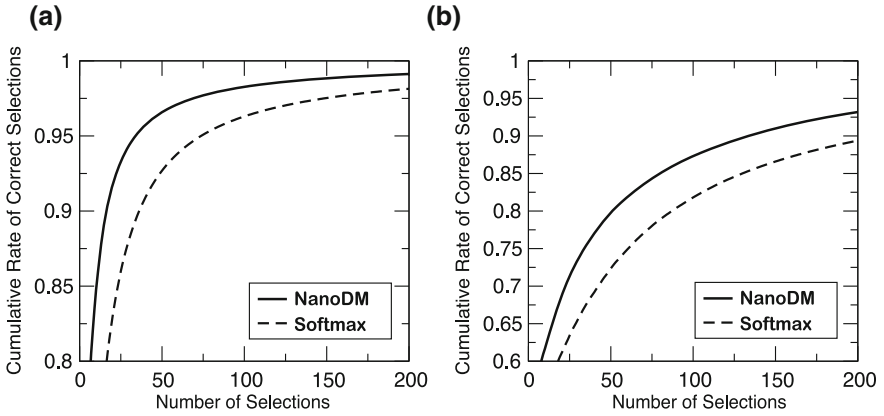


Fig. 9.10 Performance comparison between a simulation model of a nanophotonic decision maker, NanoDM (*solid lines*), and the well-known Softmax algorithm (*broken lines*) [21]. **a** Results for slot machine reward probabilities of $P_A = 0.2$ and $P_B = 0.8$. The cumulative rate of correct selections for NanoDM with fixed parameter $D = 50$ and the Softmax algorithm with optimized parameter $\tau = 0.40$ are shown. **b** Results for slot machine reward probabilities of $P_A = 0.4$ and $P_B = 0.6$. The cumulative rate of correct selections for NanoDM with fixed parameter $D = 50$ and the Softmax algorithm with optimized parameter $\tau = 0.25$ are shown

In this study we dealt with restricted problems, namely, $P_A + P_B = 1$. However, general problems can also be solved by the extended NanoDM although the IA dynamics becomes a bit more complicated.

9.5 Discussion and Conclusion

In this paper, we reviewed NanoPS and NanoDM, which are models that use photoexcitation transfer phenomena in QDs systems for solving SAT and BP, respectively. NanoPS and NanoDM outperformed conventional algorithms. In particular, the advantage of NanoPS over WalkSAT increased as the problem-size N for SAT increased. At each iteration, WalkSAT flips a single state without implementing any interaction among the variables. In contrast, NanoPS updates at most $2N$ states through a large number of interactions among the QDs, which exchange information on stimulated experiences via the bounceback control dynamics. This difference in the number of interactions might produce a large difference in the computational power [1]. The computational power of NanoPS, therefore, emerges from its "concurrent" nature, in which a large number of search processes run simultaneously in a spatially correlated manner, interfering with each other through the bounceback control dynamics.

An important issue that we should address to implement NanoPS and NanoDM experimentally is the means of introducing bounceback control dynamics and the

intensity adjuster. An external approach would be to use some external control unit such as a combination of a PC with light projection equipment. However, the external control unit may impose additional energy costs and limit the processing speed of our paradigm. On the other hand, an internal approach could implement the control dynamics using additional QDs without the need for an external control unit. It may be possible to embed the counterpart of the external control unit in the arrangement of QDs, because the bounceback rules are expressed by combining elementary logical operations, and these logical operations have already been implemented experimentally in several QD systems [26].

Our nanophotonic computing paradigm is fundamentally different from conventional optical computing or optical signal processing, which are limited by the properties of propagating light. Our paradigm also differs from the current quantum computing paradigm, which exploits a superposition of all possible states to produce a correct solution. This is because our paradigm exploits both coherent and dissipative processes. In fact, optical-near-field-mediated photoexcitation transfer is a coherent process, suggesting that an exciton could be transferred to all possible destination QD_L's via a resonant energy level, but such coherent interaction produces a unidirectional transfer by an energy dissipation process occurring in QD_L. An advantage of our paradigm is that photoexcitation transfer is 10^4 times more energy efficient than that in conventional electrically wired bit-flip circuits [29]. Furthermore, our SAT and BP solvers have great potential for wide applications such as artificial intelligence, information security, bioinformatics, and information communication technologies. Putting these facts together, our nanophotonic computing paradigm will contribute to developing low-energy-use highly versatile computers.

References

1. M. Conrad, Commun. ACM **28**, 464–480 (1985)
2. M. Dorigo, V. Maniezzo, A. Colorni, IEEE Trans. Syst. Man Cybernet. B **26**, 29–41 (1996)
3. E. Bonabeau, M. Dorigo, G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems* (Oxford University Press, New York, 1999)
4. G. Rozenberg, T. Back, J. Kok (eds.), *Handbook of Natural Computing* (Springer, Netherlands, 2012)
5. T. Nakagaki, H. Yamada, A. Toth, Nature **407**, 470 (2000)
6. A. Tero, S. Takagi, T. Saigusa, K. Ito, D.P. Bebber, M.D. Fricker, K. Yumiki, R. Kobayashi, T. Nakagaki, Science **327**(5964), 439–442 (2010)
7. M. Aono, Y.-P. Gunji, BioSystems **71**, 257–287 (2003)
8. M. Aono, Hara, K. Aihara, Commun. ACM **50**(9), 69–72 (2007)
9. M. Aono, Y. Hirata, M. Hara, K. Aihara, New Generat Comput **27**, 129–157 (2009)
10. A. Takamatsu, R. Tanaka, H. Yamada, T. Nakagaki, T. Fujii, I. Endo, Phys. Rev. Lett. **87**(7), 078102 (2001)
11. S. Takagi, T. Ueda, Physica D **237**, 420–427 (2008)
12. M. Aono, S.-J. Kim, L. Zhu, M. Wakabayashi, M. Hara, Submitted (2013)
13. M. Aono, S.-J. Kim, L. Zhu, M. Naruse, M. Ohtsu, H. Hori, M. Hara, in *Proceedings of 2012 International Conference on Nonlinear Theory and its Applications*, Majorca, 22–26 October 2012, pp. 586–589

14. S.-J. Kim, M. Aono, Hara. BioSystems **101**, 29–36 (2010)
15. S.-J. Kim, E. Nameda, M. Aono, M. Hara, in *Proceedings of 2011 International Conference on Nonlinear Theory and its Applications*, Kobe, 4–7 September 2011, pp. 176–179
16. S.-J. Kim, M. Aono, E. Nameda, M. Hara, Technical Report of IEICE (CCS-2011-025) [in Japanese], 2011, pp. 36–41
17. S.-J. Kim, M. Aono, E. Nameda, M. Hara, in *Proceedings of 2012 International Conference on Nonlinear Theory and its Applications*, Majorca, 22–26 October 2012, pp. 590–593
18. M. Naruse, T. Miyazaki, T. Kawazoe, S. Sangu, K. Kobayashi, F. Kubota, M. Ohtsu, IEICE Trans. Electron. **E88-C**, 1817–1823 (2005)
19. M. Naruse, M. Aono, S.-J. Kim, T. Kawazoe, W. Nomura, H. Hori, M.M. Hara, Ohtsu. Phys. Rev. B **86**, 125407 (2012)
20. M. Aono, M. Naruse, S.-J. Kim, M. Wakabayashi, H. Hori, M. Ohtsu, M. Hara, Langmuir 29, 7557–7564 (2013)
21. S.-J. Kim, M. Naruse, M. Aono, M. Ohtsu, M. Hara, Sci. Rep. **3**, 2370 (2013). doi:10.1038/srep02370
22. M. Ohtsu, T. Kawazoe, T. Yatsui, M. Naruse, IEEE J. Sel. Top. Quant. Electron **14**, 1404–1417 (2009)
23. T. Kawazoe, K. Kobayashi, K. Akahane, M. Naruse, N. Yamamoto, M. Ohtsu, Appl. Phys. B **84**, 243–246 (2006)
24. T. Yatsui, S. Sangu, T. Kawazoe, M. Ohtsu, S.J. An, J. Yoo, G.-C. Yi, Appl. Phys. Lett. **90**(223110), 1–3 (2007)
25. W. Nomura, T. Yatsui, T. Kawazoe, M. Naruse, M. Ohtsu, Appl. Phys. B **100**, 181–187 (2010)
26. T. Kawazoe, M. Ohtsu, S. Aso, Y. Sawado, Y. Hosoda, K. Yoshizawa, K. Akahane, N. Yamamoto, M. Naruse, Appl. Phys. B **103**, 537–546 (2011)
27. K. Akahane, N. Yamamoto, M. Tsuchiya, Appl. Phys. Lett. **93**, 041121 (2008)
28. C. Pistol, C. Dwyer, A.R. Lebeck, IEEE Micro. **28**, 7–18 (2008)
29. M. Naruse, H. Hori, K. Kobayashi, P. Holmström, L. Thylén, M. Ohtsu, Opt. Express **18**, A544–A553 (2010)
30. M. Ohtsu, K. Kobayashi, T. Kawazoe, T. Yatsui, M. Naruse, *M., Principles of Nanophotonics* (Taylor and Francis, Boca Raton, 2008)
31. T. Itoh, M. Furumiya, T. Ikehara, C. Gourdon, Solid State Commun. **73**, 2710–274 (1990)
32. M.R. Garey, D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness* (W. H. Freeman and co., New York, 1979)
33. H.J. Carmichael, *Statistical Methods in Quantum Optics 1* (Springer, Berlin, 1999)
34. U. Schöning, in *Proceedings of 40th Symposium on Foundations of Computer Science*, 1999, pp. 410–414
35. K. Iwama, S. Tamaki, in *Proceedings 15th Symposium on Discrete Algorithms*, 2004, p. 328
36. http://www.cs.ubc.ca/ hoos/SATLIB/benchm.html
37. S. Kirkpatrick, B. Selman, Science **264**, 1297–1301 (1994)
38. R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, L. Troyansky, Nature **400**, 133–137 (1999)
39. H. Robbins, Bull. Amer. Math. Soc. **58**, 527–536 (1952)
40. W. Thompson, Biometrika **25**, 285–294 (1933)
41. J. Gittins, D. Jones, A dynamic allocation index for the sequential design of experiments, in *Progress in Statistics*, ed. by J. Gans (North Holland, Amsterdam, 1974) pp. 241–266
42. J. Gittins, J.R. Stat, Soc. B **41**, 148–177 (1979)
43. R. Agrawal, Adv. Appl. Prob. **27**, 1054–1078 (1995)
44. P. Auer, N. Cesa-Bianchi, P. Fischer, Mach. Learn. **47**, 235–256 (2002)
45. L. Kocsis, C. Szepesvári, in *ECML2006, Lecture notes in artificial intelligence*, vol. 4212 (Springer, Berlin, 2006), pp. 282–293
46. S. Gelly, Y. Wang, R. Munos, O. Teytaud, O. RR-6062-INRIA, pp. 1–19 (2006)
47. N. Daw, J. O'Doherty, P. Dayan, B. Seymour, R. Dolan, Nature **441**, 876–879 (2006)

Index

M. Naruse (ed.), *Nanophotonic Information Physics*,
Nano-Optics and Nanophotonics, DOI: 10.1007/978-3-642-40224-1,