



(51) International Patent Classification:

C12Q 1/00 (2006.01) *G01N 27/26* (2006.01)
C12Q 1/02 (2006.01) *G01N 27/30* (2006.01)

P.O. Box 910657, San Diego, California 92191 (US). **COSTA, Gina**; P.O. Box 910657, San Diego, California 92191 (US).

(21) International Application Number:

PCT/US2018/048873

(74) **Agent: PAPPALARDO, Paul A.**; Snell & Wilmer L.L.P., 400 E. Van Buren St., Phoenix, Arizona 85004-2202 (US).

(22) International Filing Date:

30 August 2018 (30.08.2018)

(81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/551,977 30 August 2017 (30.08.2017) US

(71) **Applicant: ROSWELL BIOTECHNOLOGIES, INC.** [US/US]; P.O. Box 910657, San Diego, California 92191 (US).

(72) **Inventors: MERRIMAN, Barry**; P.O. Box 910657, San Diego, California 92191 (US). **GEISER, Tim**; P.O. Box 910657, San Diego, California 92191 (US). **MOLA, Paul**;

(84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,

(54) **Title:** PROGRESSIVE ENZYME MOLECULAR ELECTRONIC SENSORS FOR DNA DATA STORAGE

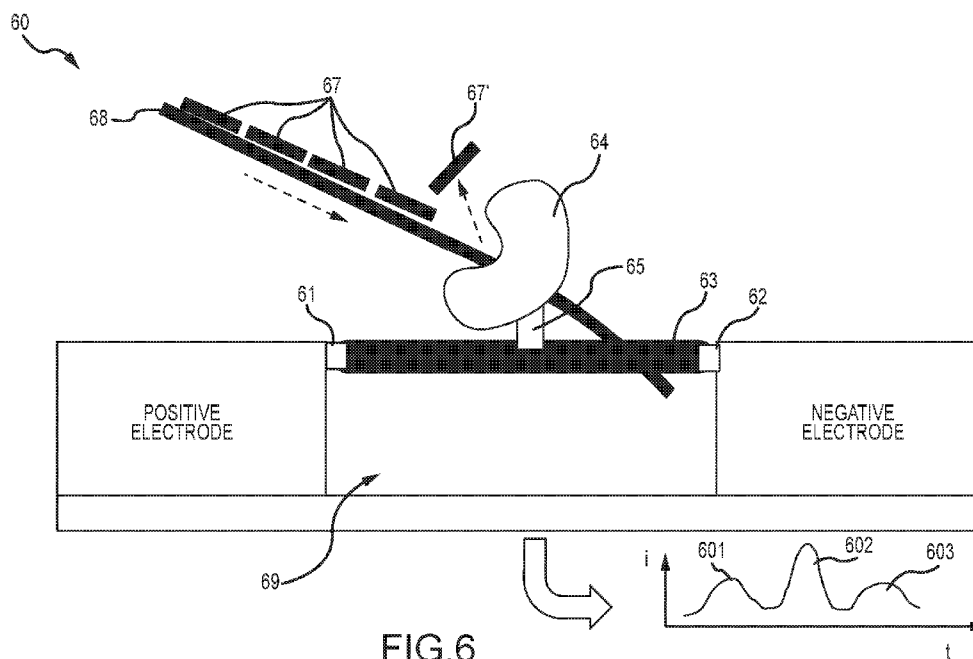


FIG. 6

(57) **Abstract:** A progressive enzyme molecular sensor for use in a DNA data storage system is disclosed that can extract digital information suitably encoded into a synthetic DNA molecule. In various aspects, such sensors are provided in a high-density chip-based format that can provide the high throughput, low-cost and fast data extraction capability required for large scale DNA data storage systems. The sensor for reading the digital data stored in DNA molecules processes individual encoded DNA molecules directly, eliminating the need for complicated sample preparation such as making copies of DNA or clonal populations of such molecules.



UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

PROCESSIVE ENZYME MOLECULAR ELECTRONIC SENSORS FOR DNA DATA STORAGE

Assignee: ROSWELL BIOTECHNOLOGIES, INC.
5 **Inventors:** Barry Merriman, Tim Geiser, Paul Mola, Gina Costa

Cross-Reference to Related Applications

This application claims priority to and the benefit of U.S. Provisional Patent Application Serial No. 62/551,977 filed August 30, 2017, entitled “Processive Enzyme
10 Molecular Electronic Sensors for DNA Data Storage,” the disclosure of which is incorporated herein by reference in its entirety for all purposes.

Field

The present disclosure generally relates to electronic data storage and retrieval, and
15 more particularly to DNA information and retrieval systems comprising molecular electronic sensors for reading information stored as DNA molecules.

Background

The advent of digital computing in the 20th Century created the need for archival
20 storage of large amounts of digital or binary data. Archival storage is intended to house large amounts of data for long periods of time, e.g., years, decades or longer, in a way that is very low cost, and that supports the rare need to re-access the data. Although an archival storage system may feature the ability to hold unlimited amounts of data at very low cost, such as through a physical storage medium able to remain dormant for long periods of time, the data
25 writing and recovery processes in such a system can be relatively slow or otherwise costly. The dominant forms of archival digital data storage developed to date include magnetic tape, and, more recently, compact optical disc (CD). However, as data production grows, there is a need for even higher density, lower cost, and longer lasting archival digital data storage systems.

30 It has been observed that in biology, the genomic DNA of a living organism functions as a form of digital information archival storage. On the timescale of the existence of a species, which may extend for thousands to millions of years, the genomic DNA in effect stores the genetic biological information that defines the species. The complex enzymatic, biochemical processes embodied in the biology, reproduction and survival of the species
35 provide the means of writing, reading and maintaining this information archive. This

observation has motivated the idea that perhaps the fundamental information storage capacity of DNA could be harnessed as the basis for high density, long duration archival storage of more general forms of digital information.

5 What makes DNA attractive for information storage is the extremely high information density resulting from molecular scale storage of information. In theory for example, all human-produced digital information recorded to date, estimated to be approximately 1 ZB (Zettabyte, 10^{21} bytes), could be recorded in less than 10^{22} DNA bases, or $1/60^{\text{th}}$ of a mole of DNA bases, which would have a mass of just 10 grams. In addition to high data density, DNA is also a very stable molecule, which can readily last for thousands of years without
10 substantial damage, and which could potentially last far longer, for tens of thousands of years, or even millions of years, such as observed naturally with DNA frozen in permafrost or encased in amber.

Summary

15 In various embodiments of the present disclosure, a data reader for use in a DNA data storage system is provided. In particular, a sensor is disclosed that can extract the digital information synthetically encoded into a single DNA molecule. As disclosed herein, sensors for use as DNA encoded data readers are processive enzyme molecular sensors. In various aspects, a plurality of such sensors are provided in an array in a high-density chip-based
20 format that can provide the high throughput, low-cost and fast data extraction capability required for large scale DNA data storage systems. In various embodiments, the sensor for reading the digital data stored in DNA molecules processes individual encoded DNA molecules directly, so that there is no need for complicated sample preparation such as making copies of DNA or clonal populations of such molecules.

25 In various aspects, the fundamental time required to extract information encoded in a DNA molecule is short, on the order of seconds, which fundamentally enables short turn-around times for data recovery. Information extraction from a DNA molecule is at the rate of a processive enzyme acting on DNA, which can be very fast, as short as fractions of a second. In various aspects, information can be encoded in native DNA, without modifications, which
30 can be replicated, copied or amplified as needed. Further, simple means to enhance the signal-to-noise ratio are disclosed. The sensor can be deployed in a highly scalable, low cost, CMOS chip format, providing for efficient mass manufacturing, and low cost systems and instruments, and overall low costs for reading digital data stored in DNA. The systems and instruments required to read Exabyte-scale digital data from DNA data can be highly

compact and energy efficient, to support practical, robust deployment both locally at on-site data centers, and for highly scalable cloud-based archival data storage services. Sensors for information extraction herein provide for systems that read data stored in DNA that can exceed the performance, in speed, throughput and cost, of the respective reader systems for data archived in conventional archival storage formats such as magnetic tape or optical discs. The present disclosure provides enabling technology for DNA digital data storage systems capable of practical Exabyte scale storage, and Zettabyte scale storage.

In various embodiments of the present disclosure, a sensor is disclosed. In various aspects the sensor comprises a processive enzyme molecular electronics sensor. In various examples the sensor comprises: a first electrode; a second electrode spaced apart from the first electrode by an electrode gap; a processive enzyme conjugated to the first and second electrodes, the processive enzyme comprising a native or genetically engineered polymerase, reverse transcriptase, helicase, exonuclease, or molecular motor for packaging of viral DNA; and a trans-impedance amplifier electrically connected to at least one of the first electrode and second electrode, the trans-impedance amplifier providing an output comprising a measurable electrical parameter; wherein the measurable electrical parameter comprises distinguishable signals corresponding to enzymatic activity of the processive enzyme.

In various aspects of a sensor, the trans-impedance amplifier provides a biasable voltage across the first and second electrodes, and the measurable electrical parameter comprises a current output.

In various aspects of a sensor, the sensor may further comprise a gate electrode capacitively coupled to the electrode gap.

In various aspects of a sensor, the trans-impedance amplifier may further provide a biasable voltage to the gate electrode when a gate electrode is utilized in the sensor.

In various aspects of a sensor, the processive enzyme is directly wired between first and second electrodes to provide a conductive pathway between the first and second electrodes, through the processive enzyme. For direct wiring, two sites on the enzyme molecule, such as two amino acid residues, may be modified to promote direct chemical bonding of each site to the first and second electrodes.

In various aspects of a sensor, the processive enzyme is conjugated to the first and second electrodes via a bridge molecule spanning the electrode gap, the bridge molecule having first and second ends, the first end conjugated to the first electrode and the second end conjugated to the second electrode, wherein the processive enzyme is conjugated to the bridge molecule.

In various aspects of a sensor, the processive enzyme is conjugated to the first and second electrodes via at least one intervening arm molecule, the arm molecule conjugated to at least one of the first and second electrodes and conjugated to the processive enzyme.

5 In various aspects of a sensor, bridge or arm molecules may comprise a double stranded DNA, a protein alpha helix, a graphene nanoribbon, a carbon nanotube, an antibody, or a Fab arm of an antibody.

In various aspects of a sensor, the first and second electrodes comprise source and drain electrodes, respectively, and wherein the measurable electrical parameter is the source-drain current between the electrodes.

10 In various aspects of a sensor, the trans-impedance amplifier further provides for voltage-biasable source, drain and gate electrodes.

In various embodiments of the present disclosure, a method of reading encoded information is disclosed. The method comprises: translocating a synthetic DNA molecule through a processive enzyme, where the processive enzyme is electrically connected in a
15 molecular sensor circuit; and generating signals in a measureable electrical parameter of the circuit, wherein the signals correspond to the encoded information.

In various aspects of a method, the processive enzyme comprises a native or genetically engineered polymerase, reverse transcriptase, helicase, exonuclease, or molecular motor for packaging of viral DNA.

20 In various aspects of a method, the encoded information comprises binary data.

In various aspects of a method, the synthetic DNA molecule comprises a DNA template strand with distinguishable signaling features bonded thereon.

In various aspects of a method, the distinguishable signaling features comprise a sequence of oligonucleotides.

25 In various aspects of a method, the oligonucleotides are bonded to the DNA template strand by complementary base pairing and the oligonucleotides are displaced from the DNA template strand by the processive enzyme as the synthetic DNA molecule processively translocates through the processive enzyme, encountering the oligonucleotides. In these examples, a signal in the measureable electrical parameter of the circuit corresponds to
30 displacement of an oligonucleotide from the DNA template strand.

In various aspects of a method, each of the oligonucleotides further comprises at least one chemical group bonded thereon.

In various aspects of a method, the oligonucleotides are covalently bonded to the DNA template strand, wherein a distinguishable signal in the measurable electrical parameter

of the circuit corresponds to interaction of an oligonucleotide with the processive enzyme without displacement of the oligonucleotide from the DNA template strand.

In various aspects of a method, each of the oligonucleotides further comprises at least one chemical group bonded thereon.

5 In various aspects of a method, the distinguishable signaling features comprise chemical groups conjugated to the DNA template strand.

In various embodiments of the present disclosure, a DNA information system is disclosed. The system comprises: a synthetic DNA molecule comprising a DNA template strand and distinguishable signaling features bonded thereon, the distinguishable signaling
10 features encoding information in the synthetic DNA molecule; a buffer solution in contact with the synthetic DNA molecule, the first and second electrodes, and the processive enzyme; and a sensor capable of reading the information encoded in the synthetic DNA molecule, the sensor comprising: a first electrode; a second electrode spaced apart from the first electrode by an electrode gap; a processive enzyme conjugated to both the first and second electrodes;
15 and a trans-impedance amplifier electrically connected to at least one of the first and second electrodes, the trans-impedance amplifier providing an output comprising a measurable electrical parameter, wherein the measurable electrical parameter comprises distinguishable signals corresponding to the encoded information.

In various aspects of a system, the processive enzyme comprises a native or
20 genetically engineered polymerase, reverse transcriptase, helicase, exonuclease, or molecular motor for packaging of viral DNA.

In various aspects of a system, the encoded information comprises binary data.

In various aspects of a system, the distinguishable signaling features comprise a sequence of oligonucleotides.

25 In various aspects of a system, the oligonucleotides are bonded to the DNA template strand by complementary base pairing, or the oligonucleotides are covalently bonded to the DNA template strand. In certain aspects, combinations of base-paired oligonucleotides and covalently bound oligonucleotides encode the information in the synthetic DNA molecule. In certain aspects, base-paired oligonucleotides are displaced from the DNA template strand by
30 the processive enzyme, whereas covalently bonded oligonucleotides are not displaced when the processive enzyme encounters them.

In various aspects of a system, each of the oligonucleotides further comprises at least one chemical group bonded thereon. In other examples, at least one oligonucleotide in a

sequence of oligonucleotides bonded to the DNA template strand comprise a chemical group bonded thereon.

In various aspects of a system, the distinguishable signaling features comprise chemical groups conjugated to the DNA template strand.

5 In various aspects of a system, the system further comprises a reference electrode submerged in the buffer solution. The reference may be, for example, a Ag/AgCl electrode.

In various aspects of a system, the system further comprises a gate electrode capacitively coupled to the electrode gap.

10 In various aspects of a system, the processive enzyme is directly wired between first and second electrodes to provide a conductive pathway between first and second electrodes, through the processive enzyme.

In various aspects of a system, the processive enzyme is conjugated to the first and second electrodes via a bridge molecule spanning the electrode gap, the bridge molecule having first and second ends, the first end conjugated to the first electrode and the second end conjugated to the second electrode, wherein the processive enzyme is conjugated to the bridge molecule.

In various aspects of a system, the processive enzyme is conjugated to the first and second electrodes via at least one intervening arm molecule, the arm molecule conjugated to at least one of the first and second electrodes and conjugated to the processive enzyme.

20 In various aspects of a system, bridge or arm molecules may comprise a double stranded DNA, a protein alpha helix, a graphene nanoribbon, a carbon nanotube, an antibody, or a Fab arm of an antibody.

In various aspects of a system, the system further comprises a CMOS sensor array chip comprising an array of the sensors and supporting pixel circuitry that performs measurement of the measurable electrical parameter.

25 In various aspects of a system, the system comprises at least two of the CMOS sensor array chips; an electronic hardware system for controlling and managing electrical inputs and data outputs of the chips; a fluidic system for introducing the synthetic DNA molecule in the buffer solution to the chips; and a signal processing and data recording system for capturing the distinguishable signals and for converting the distinguishable signals back to the information.

30 In various aspects of a system, the synthetic DNA molecule comprises a circular, hairpin, or tandem repeat architecture that allows repeat reading of the information encoded in the synthetic DNA molecule.

Brief Description of the Drawings

The subject matter of the present disclosure is particularly pointed out and distinctly claimed in the concluding portion of the specification. A more complete understanding of the present disclosure, however, may best be obtained by referring to the detailed description and claims when considered in connection with the drawing figures:

FIG. 1-A shows an embodiment of a molecular electronic sensing circuit, in which a molecule completes an electrical circuit, an electrical circuit parameter is measured versus time for signals, where the recorded signals correspond to interactions of the molecule with interacting molecules in the surrounding environment;

FIG. 1-B shows an embodiment of a processive enzyme molecular electronics sensor that can be used to read data encoded into synthetic DNA molecules. The sensor produces distinguishable signals corresponding to distinct signaling features structurally present on a template DNA molecule; such feature elements can be used to encode information into synthetic template DNA molecules, which can in turn be read via the sensor;

FIG. 2 shows representative processive enzymes that can be used in a molecular sensor to act on various forms of DNA, including specific examples of a polymerase, a helicase, an exonuclease, and a DNA packaging molecular motor. Illustrated are: Klenow + DNA, Protein DataBase (PDB) structure ID 1KLN; Exonuclease I E Coli, PDB ID 1FXX; PDB ID 1G8Y; the crystal structure of the hexameric replicative helicase REPA; PDB ID 3EZK, Bacteriophage T4 gp17 motor assembly based on crystal structures and cryo-EM reconstructions; PDB ID 2WWY; and the structure of human RECQ-like helicase in complex with a DNA substrate;

FIG. 3-A shows an embodiment of a processive enzyme molecular sensor in the process of reading DNA, in which groups bound to the template DNA are displaced by the action of the processive enzyme, with the displacement events resulting in distinguishable signal features;

FIG. 3-B shows an embodiment of a processive enzyme molecular sensor in the process of reading DNA, in which the template DNA has modifying groups attached, and as the processive enzyme translocates past these features, it results in the production of distinguishable signal features;

FIG. 4-A shows an embodiment of a processive enzyme molecular sensor comprising a processive enzyme molecule conjugated to a bridge molecule that spans the gap between two electrodes, conjugated to each electrode;

FIG. 4-B shows an embodiment of a processive enzyme molecular sensor comprising a processive enzyme molecule wired directly into the current path of the sensor by way of two arm molecules providing the connections to the electrodes;

FIG. 4-C shows an embodiment of a processive enzyme molecular sensor, wherein
5 the processive enzyme molecule is directly conjugated to the electrodes without any intervening arm molecules or a bridge molecule in the circuit;

FIG. 5 illustrates embodiments of a processive enzyme molecular sensor, wherein the processive enzyme comprises a polymerase having strand displacing activity;

FIG. 6 illustrates embodiments of a processive enzyme molecular sensor, wherein the
10 processive enzyme comprises a helicase having strand displacing activity;

FIG. 7 illustrates embodiments of a processive enzyme molecular sensor, wherein the processive enzyme comprises a DNA packaging molecular motor having the ability to displace bound oligonucleotides sufficiently sterically bulky so as not to fit through the motor inlet;

FIG. 8 illustrates embodiments of a processive enzyme molecular sensor, wherein the
15 processive enzyme comprises an exonuclease having strand displacing activity;

FIG. 9 illustrates embodiments of a processive enzyme molecular sensor, wherein the processive enzyme comprises a helicase;

FIG. 10 illustrates embodiments of a processive enzyme molecular sensor, wherein
20 the processive enzyme comprises a DNA packaging molecular motor;

FIG. 11 illustrates embodiments of a processive enzyme molecular sensor, wherein the processive enzyme comprises an exonuclease;

FIG. 12-A shows the detailed protein structure of one specific processive enzyme molecule, a polymerase comprising the Klenow Fragment of E. Coli. Polymerase I, with (A)
25 and without (B) a DNA substrate interacting with the polymerase;

FIG. 12-B shows the detailed protein structure of one specific processive enzyme molecule, a helicase comprising the human RECQ-like DNA helicase while interacting with a DNA substrate molecule;

FIG. 13-A shows embodiments of a processive enzyme molecular electronic sensor,
30 wherein the polymerase processive enzyme molecule is conjugated to a bridge molecule bonded between the electrodes;

FIG. 13-B shows an embodiment of the processive enzyme molecular electronic sensor of FIG. 13-A, wherein the bridge molecule comprises a double stranded DNA, the polymerase-bridge conjugation comprises biotin-streptavidin binding, and the electrodes

comprise gold-on-chromium to support thiol-gold binding of first and second ends of the bridge molecule to the electrodes;

FIG. 13-C shows embodiments of a processive enzyme molecular electronic sensor, wherein the polymerase processive enzyme is conjugated directly into the current path by use
5 of two arm molecules between the enzyme and each of the two electrodes;

FIG. 13-D shows embodiments of a processive enzyme molecular electronic sensor, wherein the polymerase processive enzyme is conjugated directly into the current path, and directly to the metal electrodes, without use of arm or bridge molecules;

FIG. 13-E shows embodiments of a processive enzyme molecular electronic sensor,
10 wherein the processive enzyme comprises a helicase (human RECQ-like helicase) and is conjugated to a bridge molecule that spans the gap between the spaced-apart electrodes;

FIG. 14-A shows the physical structure of various embodiments of different DNA data encoding template molecules, wherein the signaling features are: (A) bound DNA oligonucleotides, (B) bound DNA oligonucleotides further comprising additional modifying
15 groups, (C) single stranded DNA further comprising modifying groups, and (D) double stranded DNA further comprising modifying groups;

FIG. 14-B shows template structures (strand architectures) that allow a data payload from a single DNA molecule to be read multiple times by the same processive enzyme molecular sensor;

FIG. 15 shows various embodiments of the logical structure for DNA data storage molecules, comprising handling adapters, a start site for the processive enzyme to engage, buffer segments, and a data payload segment encoding the primary binary data payload;

FIG. 16 shows examples of Binary Data Encoding Schemes (“BES”) that can be used to encode digital information into the DNA data payload segment of a synthetic DNA
25 molecule, for later reading by a processive enzyme molecular electronic sensor;

FIG. 17-A shows an embodiment of a fabrication stack usable to put a plurality of DNA reader sensors on a chip for massively parallel, low-cost deployment of a DNA reading system;

FIG. 17-B shows an embodiment of CMOS chip and pixel architecture for a chip-
30 based array of processive enzyme molecular sensors. In this embodiment, the sensor is formatted into scalable CMOS chip sensor array architecture, with molecular electronic sensor and source, gate and drain terminals;

FIG. 18 shows an embodiment of a circuit schematic and the resulting measurement for the pixel circuit of FIG. 17-B, comprising sensing of a 10 pA current, for one possible choice of circuit parameters (transistor properties, resistors and capacitors);

FIG. 19 shows the completed, annotated chip design, and a drawing of an optical microscope image of the fabricated chip, for an embodiment of the pixel array chip of FIG. 17-B having an array of 256 pixels;

FIG. 20 shows a drawing of an electron microscope image of the fabricated chip of FIG. 19, including insets of the nano-electrode with a polymerase processive enzyme molecular complex in place;

FIG. 21 shows a schematic of a complete system for reading DNA data with chip-based DNA reader comprising processive enzyme molecular electronic sensors;

FIG. 22 shows a schematic of a cloud-based DNA data archival storage system, in which a multiplicity of the DNA reading system of FIG. 21 are aggregated to provide the data reader server;

FIG. 23 shows an alternate embodiment of a DNA data reader sensor in which a processive enzyme is complexed with a different electronic sensor configuration than in FIG. 1, here a nanopore ion current sensor, and which produces distinguishable signal features in the nanopore ion current when processing a DNA template molecule having signaling features;

FIG. 24 shows an embodiment of DNA data reader sensor of FIG. 23, in which the Helicase is directly conjugated to the nanopore, and in which the signaling groups are oligonucleotides bound to the DNA template, which further carry groups that alter the current in the nanopore sensor as they are translated past the pore by the action of the Helicase;

FIG. 25 shows the concept of a DNA data reader sensor which a processive enzyme molecule is complexed with a carbon nanotube molecular wire spanning positive and negative electrodes, and produces distinguishable signal features in the measured current passing through the carbon nanotube;

FIG. 26 shows an embodiment of a Zero Mode Waveguide sensor complexed with a processive enzyme molecule, shown in cross section, which produces distinguishable optical signals corresponding to DNA features, here due to dye molecules attached to oligonucleotides bound to the template DNA; and

FIG. 27 shows embodiments of a processive enzyme molecular electronic sensor, capable of reading information stored in a polymer other than DNA, comprising a lysozyme

enzyme that produces distinguishable signals from signaling groups attached to a peptidoglycan polymer that is digested by the enzyme.

Detailed Description

5 In various embodiments, a molecular electronics sensor comprising a processive enzyme is disclosed that extracts information from DNA molecules by reading digital data stored as DNA. The present disclosure further provides a means of deploying such sensors in a chip-based format, and a resulting data reading system that supports such a chip-based sensor device. Reading information encoded into DNA molecules by use of the processive
10 enzyme molecular sensors of the present disclosure is fundamental to overall methods and systems for DNA data storage. Various aspects of these methods and systems for DNA data storage, including aspects of molecular sensors for reading DNA molecules encoded with digital information, is disclosed in PCT Application Serial No. PCT/US2018/013140, filed January 10, 2018 and entitled “METHODS AND SYSTEMS FOR DNA DATA
15 STORAGE,” the disclosure of which is incorporated herein by reference in its entirety for all purposes.

Definitions

As used herein, the term “DNA” may refer not only to the biological DNA molecule,
20 but also to fully synthetic versions, made by the methods of synthetic chemistry, such as nucleotide phosphoramidite chemistry, or by serial ligation of DNA oligomers, and also to forms made with chemical modifications present on the bases, sugar, or backbone, of which many are known to those skilled in nucleic acid biochemistry, including methylated bases, adenylated bases, other epigenetically marked bases, or also including non-standard or
25 universal bases, such as inosine or 3-nitropyrrole, or other nucleotide analogues, or ribobases, or abasic sites, or damaged sites, and also including such DNA analogues as Peptide Nucleic Acids (PNA), Locked Nucleic Acids (LNA), Xeno Nucleic Acids (XNA) (a family of sugar-modified forms of DNA, including Hexitol Nucleic Acid (HNA)), Glycol Nucleic Acid (GNA), etc., and also including the biochemically similar RNA molecule along with
30 synthetic RNA and modified forms of RNA. All these biochemically closely related forms are implied by the use of the term DNA, in the context of referring to the data storage molecule used in a DNA storage system, including a template single strand, a single strand with oligomers bound thereon, double stranded DNA, and double strands with bound groups such as groups to modify various bases. In addition, as used herein, the term “DNA” may

refer to the single stranded forms of such molecules, as well as double helix or double-stranded forms, including hybrid duplex forms, including forms that containing mismatched or non-standard base pairings, or non-standard helical forms such as triplex forms, as well as molecules that are partially double stranded, such as a single-stranded DNA bound to an oligonucleotide primer, or a molecule with a hairpin secondary structure. In various
5 embodiments, “DNA” refers to a molecule comprising a single-stranded DNA component having bound oligonucleotide segments and/or perturbing groups that can act as the substrate for a processive enzyme to process, and in doing so, generate distinguishable signals in a monitored electrical parameter of a molecular sensor comprising the processive enzyme.

10 DNA sequences as written herein, such as GATTACA (SEQ ID NO: 1), refer to DNA in the 5' to 3' orientation, unless specified otherwise. For example, GATTACA (SEQ ID NO: 1) as written herein represents the single stranded DNA molecule 5'-G-A-T-T-A-C-A-3' (SEQ ID NO: 1). In general, the convention used herein follows the standard convention for written DNA sequences used in the field of molecular biology.

15 As used herein, the term “dNTP” may refer not only to the standard, naturally occurring nucleoside triphosphates used in biosynthesis of DNA—i.e. dATP, dCTP, dGTP, and dTTP—but also to natural or synthetic analogues or modified forms of these, include those that carry base modifications, sugar modifications, or phosphate group modifications, such as an alpha-thiol modification or gamma phosphate modifications, or the tetra-, penta-,
20 hexa- or longer phosphate chain forms, or possibly with additional groups conjugated to any of the phosphates, particularly the beta, gamma or higher order phosphates in the chain. In general, as used herein, “dNTP” may mean any nucleoside triphosphate analogue or modified form that can be incorporated by a polymerase enzyme as it extends a primer, or that would enter the active pocket of such an enzyme and engage transiently as a trial candidate for
25 incorporation.

As used herein, the term “oligonucleotide” or “binding oligonucleotide” refers to a short segment of DNA, or analog forms described above, having a length in the range of 3—100 bases, or 5—40 bases, or 10—30 bases, which can hybridize with complementary sequence contain in a template strand. Such hybridization may be through perfect Watson-
30 Crock base-pairing matches, or may involve mismatches or nonstandard base pairings.

As used herein, the term hybridization refers to the process of complementary strand binding as it occurs in the DNA double helix, or in similar duplex formation for DNA analogs.

As used herein, “buffer,” “buffer solution” and “reagent solution” refers to a solution which provides the environment in which the processive enzyme sensor can operate and produce signals from supplied DNA templates. In various embodiments, the solution is an aqueous solution, which may comprise dissolved, suspended or emulsified components such as salts, pH buffers, divalent cations, surfactants, blocking agents, solvents, template primer
5 oligonucleotides, other proteins that complex with the polymerase, and also possibly including the polymerase substrates, i.e. dNTPs, analogues or modified forms of dNTPs, and DNA molecule substrates or templates.

As used herein, “binary data” or “digital data” refers to data encoded using the standard binary code, or a base 2 {0,1} alphabet, data encoded using a hexadecimal base 16
10 alphabet, data encoded using the base 10 {0—9} alphabet, data encoded using ASCII characters, or data encoded using any other discrete alphabet of symbols or characters in a linear encoding fashion.

As used herein, “digital data encoded format” refers to the series of binary digits, or
15 other symbolic digits or characters that come from the primary translation of the DNA sequence features used to encode information in DNA, or the equivalent logical string of such classified DNA features. In some embodiments, information to be archived as DNA may be translated into binary, or exist initially as binary data, and then this data may be further encoded with error correction and assembly information, into the format that is directly
20 translated into the code provided by the distinguishable DNA signaling features. This latter association is the primary encoding format of the information. Application of the assembly and error correction procedures is a further, secondary level of decoding, back towards recovering the source information.

As used herein, the term “signaling feature” refers to a characteristic of a data-
25 encoding DNA molecule that, when encountered and processed by the processive enzyme of a processive enzyme molecular sensor, produces a signal in a monitored electrical parameter of the sensor circuit, such as current (i). Arrangements of signaling features on a DNA molecule are used to encode information in a synthetic DNA molecule. The broader group of signaling features herein further comprise both “bound groups” capable of displacement by a
30 processive enzyme of a sensor and “perturbing groups” that are not displaced from the DNA molecule by the processive enzyme. Both types of signaling features on a DNA molecule provide distinct signals in a monitored electrical parameter of the sensor circuit when encountered by the processive enzyme of the sensor. Signaling features comprise, for example, hybridization-bound oligonucleotides, chemical groups conjugated to the DNA, or

combinations of such to achieve arrangements of features that produce distinguishable signals when processed by the sensor processive enzyme.

As used herein, a “data-encoding DNA molecule,” or “DNA data encoding molecule,” refers to a DNA molecule synthesized to encode data, such as binary information,
5 in its molecular structure, including copies of information containing DNA molecules or other DNA molecules derived from such molecules, such as complementary sequences.

As used herein, “reading data from DNA” refers to any method of measuring distinguishable events, such as electrical signals or other perturbations in a monitored electrical parameter of a circuit, which correspond to molecular features in a synthetic DNA
10 molecule that were used to encode information into the DNA molecule.

As used herein, “electrodes” refer to nano-scale electrical conductors (more simply, “nano-electrodes”), disposed in pairs and spaced apart by a nanoscale-sized electrode gap between the two electrodes in any pair of electrodes. In various embodiments, the term “electrode” may refer to a source, drain or gate. A gate electrode may be capacitively coupled
15 to the gap region, and may be a “buried gate,” “back gate,” or “side gate.” The electrodes in a pair of spaced-apart electrodes may be referred to specifically (and labeled as such in various drawing figures) as the “source” and “drain” electrodes, “positive” and “negative” electrodes, or “first” and “second” electrodes. Whenever electrodes in any of the drawing figures herein are labeled “positive electrode” and “negative electrode,” it should be understood the polarity
20 indicated may be reversed, (i.e., the labels of these two elements in the drawings can be reversed), unless indicated otherwise, (such as an embodiment where electrons may be flowing to the negative electrode). Nano-scale electrodes in a pair of electrodes are spaced apart by an electrode gap measuring about 1 nm to 100 nm, and may have other critical dimensions, such as their width, height, and length, also in this same nanoscale range. Such
25 nano-electrodes may be composed of a variety of materials that provide conductivity and mechanical stability. They may be comprised of metals, or semiconductors, for example, or of a combination of such materials. Pairs of spaced-apart electrodes may be disposed on a substrate by nano-scale lithographic techniques.

As used herein, the term “enzyme” refers to any molecule or molecular complex that
30 acts on a substrate molecule to alter its state. Such enzymes are often proteins or comprise a protein component.

As used herein, a “processive enzyme” refers to any enzyme having “processivity,” which is the ability to catalyze consecutive reactions without the enzyme releasing its substrate. In various embodiments, a processive enzyme, as all or part of its enzymatic

activity, may process a DNA molecule by engaging the DNA molecule at one end of the molecule or at an internal initiation site and translocating along the DNA molecule. This translocative movement is, of course, relative, and it may be the substrate that moves through a processive enzyme if the enzyme is physically anchored in a molecular sensor. Processive enzymes include, but are not limited to, polymerases, helicases, exonucleases, and molecular motors for packaging virus DNA.

As used herein, a “bridge molecule” refers to a molecule bound between two spaced-apart electrodes in a pair of electrodes, to span the electrode gap there between, and complete an electrical circuit. In various embodiments, a bridge molecule has roughly the same length as an electrode gap, such as 1 nm to 100 nm, or in some cases, about 10 nm. Bridge molecules for use herein may comprise double stranded DNA, other analog DNA duplex structures, such as DNA-RNA, DNA-PNA or DNA-LNA or DNA-XNA duplex hybrids, peptides, protein alpha-helix structures, antibodies or antibody Fab domains, graphene nanoribbons or carbon nanotubes, silicon nanowires, or any other of a wide array of molecular wires or conducting molecules known to those skilled in the art of molecular electronics. A bridge molecule herein may be described as having a “first” and “second” end, such as a base at or near the 3’ end and a base at or near the 5’ end of a DNA molecule acting as a bridge molecule. For example, each end may be chemically modified such that the first end of a bridge molecule bonds to a first electrode and the second end of a bridge molecule bonds to a second electrode in a pair of spaced-apart electrodes. This nomenclature aids in visualizing a bridge molecule spanning an electrode gap and bonding to each electrode in a pair of spaced-apart electrodes. In various embodiments, the first and second ends of a bridge molecule may be chemically modified so as to provide for self-assembly between the bridge molecule and a processive enzyme and/or between the bridge molecule and one or both electrodes in a pair of electrodes.

As used herein, an “arm” molecule has many of the same characteristics as a bridge molecule, and may comprise the same chemical species, e.g., DNA, although an arm molecule may be of shorter molecular length so that it may assist bonding a processive enzyme to only one electrode in a pair of spaced-apart electrodes. In various aspects, at least two arm molecules may be used to electrically connect a processive enzyme to both electrodes in a pair of spaced-apart electrodes, thereby suspending it between electrodes without directly bonding the enzyme to either electrode.

As used herein, the term “conjugation” refers to a chemical linkage, (i.e., bond), of any type known in the chemical arts, e.g., covalent, ionic, Van der Waals, etc. The

conjugations of a processive enzyme to bridge and/or arm molecules, or conjugations between bridge or arm molecules to an electrode, may be accomplished by a diverse array of conjugation methods known to those skilled in the art of conjugation chemistry, such as biotin-avidin couplings, thiol-gold couplings, cysteine-maleimide couplings, gold binding peptides or material binding peptides, click chemistry coupling, Spy-SpyCatcher protein interaction coupling, or antibody-antigen binding (such as the FLAG peptide tag/anti-FLAG antibody system), and the like. Conjugation of a processive enzyme to each electrode in a pair of spaced-apart electrodes comprises an “electrical connection” or the “electrical wiring” of the enzyme into a circuit that includes the enzyme and the pair of electrodes. In other words, the enzyme is conjugated to each electrode in a pair of electrodes to provide a conductive pathway between the electrodes that would be otherwise be insulated from one another by the electrode gap separating them. A conductive pathway is provided by electron delocalization/movement through the chemical bonds of the enzyme, such as through C-C bonds.

Further definitions, and other aspects of molecular electronic sensors, are disclosed in PCT Application Serial No. PCT/US2018/029382, filed on April 25, 2018, and U.S. Patent No. 10,036,064, issued July 31, 2018, the disclosures of which are incorporated herein by reference in their entireties for all purposes.

FIG. 1-A shows an embodiment of a molecular electronic sensing circuit in which a molecule completes an electrical circuit and an electrical circuit parameter is measured versus time to provide a signal, wherein variations in signal reflect interactions of the molecule with other molecules in the environment. As illustrated in FIG. 1-A, a molecular electronics sensor 1 comprises a circuit in which a single sensor molecule 2, (or alternatively, a sensor complex comprising a small number of molecules), forms a completed electrical circuit by spanning the electrode gap 9 between a pair of spaced-apart nano-scale electrodes 3 and 4, comprising for example positive and negative electrodes, respectively, disposed on a support layer 5. The sensor molecule may be electrically conjugated in place to each of the electrodes by specific attachment points 6 and 7. In certain aspects, an electronic parameter 100 of the circuit is measured as the sensor molecule 2 interacts with various interacting molecules 8 to provide signals 101 in the measured electronic parameter. The measured parameter 100 may comprise current (i) passing between the electrodes and through the sensor molecule 2 versus time, with the electrical signals 101 in the measured parameter indicative of molecular interactions between the interacting molecules 8 and the sensor molecule 2, as illustrated by the plot of (i) versus (t) in FIG. 1-A.

With reference now to FIG. 1-B, a processive enzyme molecular electronics sensor 10 provides a reader for reading digital data encoded into a DNA molecule 18. In this example, a molecular complex 12 comprises a molecular bridge 13 and a processive enzyme 14 bonded thereto for reading the digital data. The single enzyme molecule 14 processively engages with a target DNA molecule 18, translocating along the DNA as it acts on it, and by doing so produces electrical signals 102 in the measured electronic parameter of the circuit as it processes along the DNA template, as shown in the plot of (i) versus (t) at the left of the figure. In various examples, the DNA template molecule 18 is augmented with “signaling features” 17, each of which results in a distinguishable electrical signal 102 in the (i) versus (t) plot when the processive enzyme 14 encounters the signaling feature. In this way, the processive enzyme sensor 10 produces a series of distinguishable electrical signals (e.g., 103 and 104 shown in the inset plot of (i) versus (t)), corresponding to the specific distinct signaling features 17 provided in predetermine patterns on the template DNA molecule 18. The inset in FIG. 1-B shows the alignment between the signaling features on the DNA molecule (indicated as “Feat. 1” and “Feat. 2” in the inset) and the distinguishable electrical signals visible in the monitored electrical parameter (indicated as 103 and 104 in the plot of (i) versus (t)). Arrangements of distinguishable signal features 17 can therefore encode information in a synthetic DNA molecule, and may be incorporated into DNA through a variety of encoding schemes, (e.g., discussed below in reference to FIG. 16).

The processive enzyme is a critical element within a processive enzyme molecular sensor that is capable of reading DNA molecules. There are many processive enzymes capable of processively engaging with DNA, such as to perform various biological functions in various organisms. The important feature here is that the enzyme translocates along the DNA molecule in the course of its function. Such enzymes typically have a specific recognition structure that may occur at one end of a DNA strand, and they engage at that site, and initiate a process in which they translocate along the DNA molecule in one direction, perhaps performing some further function as they move along the molecule. Such translocation can in some cases be entirely unidirectional or can in other cases be predominantly in one direction, but with the possibility of reverse motions. Such translocation may process the entire DNA molecule from one end to the other, or may halt or disengage before reaching the other end.

FIG. 2 illustrates non-limiting representatives of some major categories of processive enzymes that act on DNA templates. Shown from left to right are examples of a polymerase 21, an exonuclease 22, a helicase 23, a molecular motor 24 for packaging DNA, and another

helicase 25. Each of these types of processive enzymes find use in a processive enzyme molecular sensor herein, and each perform a different major function in regards to processing DNA.

The polymerase 21 illustrated in FIG. 2 is the E. Coli Klenow fragment polymerase (shown is structure 1KLN of the Protein Database (PDB)). In general, a polymerase synthesizes a complementary strand as it translates along single-stranded template DNA. Of use in the molecular sensors of the present disclosure are polymerases having strand-displacing activity, meaning that, as they translocate along single-stranded DNA they displace the complementary strand to expose the template for the strand being synthesized if they encounter a double stranded segment.

The exonuclease 22 illustrated in FIG. 2 is the E. Coli exonuclease I (PDB structure 1FXX). In general, an exonuclease digests a DNA strand as it translocates along it, one base at a time, starting from one end. Various forms act on single or double stranded DNA, and may digest one or both strands. Exonucleases may have a chemical polarity in the direction they digest the strand (i.e., 3' to 5' or 5' to 3'), and may require different types of initiation. Exonucleases may behave in different ways when they encounter various alterations in the strands they are processing, such as a change between single and double stranded forms, or the presence of chemically modified bases, or adducts, or attached groups.

One of the helicases, helicase 23, illustrated in FIG. 2 is the bacterial Helicase REPA, (PDB ID 1G8Y), which is a hexameric protein complex. The helicase 25 illustrated in FIG. 2 is the human RECQ-like DNA helicase. In general, a helicase unzips the complementary strands as it translocates along double stranded DNA. Various types of helicases may also purely translocate along single-stranded DNA, performing no other activity.

The packaging motor protein 24 illustrated in FIG. 2 is the motor protein gp17 from the T4 bacteriophage (PDB ID 3EZK). In general, a DNA packaging motor translocates a DNA strand into the capsule of a viral particle as part of the process of viral replication. Such motors may translocate along a single or double stranded DNA molecule.

FIG. 2 illustrates only non-limiting examples of enzymes that processively engage with DNA. There are many other specific proteins from these categories, as well as other forms of processive enzymes well known to those in the field of molecular biology that find use in the processive enzyme molecular sensors of the present disclosure. Further, the enzymes illustrated in FIG. 2 comprise the native forms of these enzymes. It is understood that for the processive enzyme molecular sensors of the present disclosure, native enzymes may be used, or enzymes that have been genetically modified to provide conjugation groups

at specific sites, such as to modify or augment their activity, rate of translocation, or tolerance for modifications on a DNA molecule.

FIG. 3-A illustrates various embodiments of a processive enzyme molecular electronics sensor 30 capable of reading digital information encoded and stored in DNA. The processive enzyme molecular sensor 30 comprises a molecular sensor complex 32 spanning electrode gap 39 and further comprising a bridge molecule 33 and a processive enzyme 34 bound thereto. In these embodiments, the DNA template molecule 38 comprises bound groups 37, capable of displacement from the DNA as the processive enzyme 34 translocates along the DNA molecule 38. In these examples, a displacement event, comprising displacement of a bound group 37 from the DNA molecule 38 to an unbound group 37', generates a corresponding signal feature in a monitored electrical parameter such as current (i) through the sensor circuit versus time (t). As illustrated in FIG. 3-A, the displacement events may appear as peaks 301, 302, 303 and so forth, or other distinguishable electrical signals, corresponding to each of the displacement events, in the (i) versus (t) plot. As detailed herein, there are different types of such bound groups for use herein that produce distinguishable signal features when displaced by the processive enzyme of the sensor. Arrangements of such bound groups 37 comprise encoded digital information, encoded in DNA molecules such as, for example, by the encoding schemes of FIG. 16. In various embodiments, bound groups 37 may comprise DNA oligonucleotides hybridized to single stranded DNA, or oligonucleotides further comprising a signal enhancing modification such as an attached charged group or structural group. Such constructs are illustrated in the upper two structures of FIG. 14, discussed herein.

FIG. 3-B illustrates another series of embodiments of a processive enzyme molecular electronics sensor 31 for reading digital information stored in DNA. The molecular sensors 31 comprise a molecular sensor complex 312 spanning the electrode gap 319 and further comprising a bridge molecule 313 and a processive enzyme 314 bonded thereto. In these embodiments, a DNA template molecule 318 further comprises perturbing groups 317 integrated into the molecule, wherein such groups 317 remain on the DNA template molecule as the DNA molecule translocates through the processive enzyme 314, and where in so doing the local perturbations present in the molecule result in distinguishable perturbation signals 304, 305, and 306 in the monitored electrical parameter, such as (i) versus (t). In this way, arrangements of perturbing groups integrated in DNA molecules can be used to store information in DNA molecules, such as for example by the encoding schemes of FIG. 16. In various embodiments, such perturbing groups 317 may comprise modifications to the bases

of the DNA, including modification to the nucleobase, the sugar residue, or the phosphate backbone. A very large number of such base modifications are well known to those in the field of molecular biology. Modifications may further comprise groups conjugated to modified bases, such as by any standard means of conjugation, including click chemistry
5 coupling, biotin-based coupling, N-hydroxysuccinimide (NHS) conjugation, maleimide conjugation, or any of many other methods known for conjugating groups to DNA.

The processive enzyme molecular sensor embodiments illustrated in FIGS. 3-A and 3-B comprise a molecular complex. In these drawings, the molecular complex is illustrated generically as comprising a bridge molecule (oval shape) and a processive enzyme (globular
10 shape) bonded thereto. The molecular complex is electrically connected between a pair of spaced-apart electrodes to span the electrode gap and complete an electronic circuit. The processive enzyme may be integrated as part of the molecular complex in various ways. FIGS. 4-A, 4-B, and 4-C illustrate various non-limiting embodiments for this complex and how the enzyme is integrated into an electronic circuit that comprises the molecular complex.

FIG 4-A shows an embodiment of a processive enzyme molecular sensor 40 in which
15 the processive enzyme 44 is conjugated, via a one or more conjugations points 46, to a bridge molecule 43 that spans the electrode gap 49. The bridge molecule 43 comprises first and second ends capable of attachment to each of the electrodes at attachment points 47 and 48. In this way, the bridge molecule is electrically connected to each electrode in a pair of
20 spaced-apart electrodes, such as the positive and negative electrodes shown, to span the electrode gap 49 between electrodes and complete the sensor circuit.

FIG. 4-B shows an embodiment of a processive enzyme molecular sensor 410 in which the processive enzyme 412 is conjugated directly into the current path between the electrodes, spanning the electrode gap 419, by conjugation to two separate “arm” molecules
25 413 and 414, which in turn are conjugated at one of their ends to each of the two electrodes via attachment points 416 and 417, and to bonding sites on the enzyme 412 at conjugation points 415 and 418, as shown.

FIG. 4-C shows an embodiment of a processive enzyme molecular sensor 420 in which the processive enzyme 422 is conjugated directly to the electrodes to span the
30 electrode gap 430 via the conjugation points 428 and 425, without any intervening arm or bridge molecules to assist in conjugating the enzyme to the electrodes. In this way, the processive enzyme 422 is directly wired into and functions as an element of the sensor circuit.

In general, for various embodiments of processive enzyme molecular electronic sensors, such as illustrated in FIG. 1-B through FIG. 4-C, there are many specific options for the components within the sensors. In various embodiments, the enzyme may be a native or mutant form of a polymerase, a reverse transcriptase, a helicase, an exonuclease, or a molecular motor for packaging DNA into a virus. In other embodiments, the mutated enzyme forms enable site specific conjugation of the enzyme to a bridge molecule, to one or more arm molecules, or to each of the electrodes in a pair of spaced-apart electrodes, through introduction of specific conjugation sites in the enzyme. Such conjugation sites engineered into the protein by recombinant methods or methods of synthetic biology, may in various embodiments comprise any one of a cysteine, an aldehyde tag site (e.g. the peptide motif CxPxR), a tetracysteine motif (e.g. the peptide motif CCPGCC (SEQ ID NO: 2)), and an unnatural or non-standard amino acid (NSAA) site, such as through the use of an expanded genetic code to introduce a p-acetylphenylalanine, or an unnatural crosslinkable amino acid, such as through the use of RNA- or DNA-protein cross-link using 5-bromouridine, (*see* Gott, J.M., et al., *Biochemistry*, 30 (25), 6290-6295 (1991)).

In various aspects, the bridge molecules or arm molecules may comprise double stranded DNA, other analog DNA duplex structures, such as DNA-RNA, DNA-PNA or DNA-LNA or DNA-XNA duplex hybrids, peptides, protein alpha-helix structures, antibodies or antibody Fab domains, graphene nanoribbons or carbon nanotubes, silicon nanowires, or any other of a wide array of molecular wires or conducting molecules known to those skilled in the art of molecular electronics. The conjugations of a processive enzyme to such molecules, or of such molecules to the electrodes, may be accomplished by a diverse array of conjugation methods known to those skilled in the art of conjugation chemistry, such as biotin-avidin couplings, thiol-gold couplings, cysteine-maleimide couplings, gold binding peptides or material binding peptides, click chemistry coupling, Spy-SpyCatcher protein interaction coupling, or antibody-antigen binding (such as the FLAG peptide tag/anti-FLAG antibody system) etc. Coupling of molecules to electrodes may be done via material binding peptides, or through the use of a SAM (Self-Assembling-Monolayer) or other surface derivatization on the electrode surface to present suitable functional groups for conjugation, such as azide or amine groups. The electrodes herein comprise electrically conducting nanoscale-dimensioned structures, which may comprise a metal, such as gold, silver, platinum, palladium, aluminum, chromium, or titanium, layers of such metals, such as gold on chromium, or semiconductors, such as doped silicon, or doped germanium. In various embodiments, a contact point of a first material may be disposed on a supporting second

material, such that the contact point is a site that directs chemical self-assembly of the molecular complex to the electrode. In non-limiting examples, a sensor may comprise spaced-apart pairs of titanium or platinum electrodes having a gold dot contact point deposited on each one of the electrodes to direct self-assembly of bridge molecules to the contact points such that only one bridge molecule spans each electrode gap between a pair of spaced-apart electrodes.

FIG. 5 shows embodiments of a processive enzyme molecular sensor 50 wherein the processive enzyme comprises a polymerase 54 having strand displacing activity. As shown, sensor 50 comprises processive enzyme 54 conjugated at one or more attachment points 55 to a bridge molecule 53 bonded between electrode pairs and spanning the electrode gap 59. The bridge molecule 53 comprises first and second ends functionalized to bond to the electrode pair at conjugation points 51 and 52. In this case, the processive translocation occurs as the polymerase 54 extends the 3' end of the primed strand 58, and synthesizes the complementary strand 58' of the underlying template strata. This takes place in a suitable buffer that also supplies dNTPs 56 for the synthesis of the complementary strand, as indicated. In the embodiment shown in FIG. 5, the signaling features 57 are DNA oligonucleotides hybridized to the template strand 58. The strand displacing polymerase 54 will displace and remove these oligonucleotides (as exemplified by the displacement of oligonucleotide 57') as it translocates in the course of synthesis. Each strand displacement event may generate a signal in the measured current of the circuit, as indicated by perturbations 501, 502 and 503 in the (i) versus (t) plot. The hybridizing oligonucleotides 57 may be DNA or DNA analogs, and may have further groups attached to enhance signaling, such as described further below in the context of FIG. 14.

FIG. 6 shows further embodiments of a processive enzyme molecular sensor 60 wherein the processive enzyme 64 comprises a helicase also having strand displacing activity. As shown, sensor 60 comprises processive enzyme 64 conjugated at one or more attachment points 65 to a bridge molecule 63 bonded between electrode pairs and spanning the electrode gap 69. The bridge molecule 63 comprises first and second ends functionalized to bond to the electrode pair at conjugation points 61 and 62. In this case, the signaling features 67 are DNA oligonucleotides hybridized to the template strand 68. The strand displacing helicase 64 will displace and remove these oligonucleotides (such as illustrated by displacement of oligonucleotide 67') as it translocates. Each strand displacement event may generate a signal in the measured current of the circuit, as indicated by the perturbations 601, 602, and 603 in the (i) versus (t) plot. The hybridizing oligonucleotides 67 in question

comprise DNA or DNA analogs, and may have further groups attached to enhance signaling, such as described further below in the context of FIG. 14.

FIG. 7 shows further embodiments of a processive enzyme molecular sensor 70 wherein the processive enzyme 74 comprises a DNA packaging molecular motor also having the ability to displace a bound oligonucleotide 77' that is sufficiently sterically bulky to not fit through the motor inlet. As shown, sensor 70 comprises processive enzyme 74 conjugated at one or more attachment points 75 to a bridge molecule 73 bonded between electrode pairs and spanning the electrode gap 79. The bridge molecule 73 comprises first and second ends functionalized to bond to the electrode pair at conjugation points 71 and 72. In this case, the signaling features comprise DNA oligonucleotides 77 hybridized to the template strand 78. The strand displacing motor 74 will displace and remove these oligonucleotides (such as 74' thus displaced) as it translocates the DNA. Each strand displacement event may generate a signal in the measured current of the circuit, as indicated by perturbations 701, 702, and 703 in the (i) versus (t) plot. The hybridizing oligonucleotides 77 in question may comprise DNA or DNA analogs, and may have further groups attached to enhance signaling, such as described further below in the context of FIG. 14. In these embodiments where the processive enzyme comprises a molecular motor, the oligonucleotides 77 may include a sterically hindering group that makes them too large to fit through the motor inlet, such as a polyethylene glycol (PEG) group, an Avidin protein group, or any other bulky group that can be readily attached to a DNA oligonucleotide. Many such groups are known and readily available to those skilled in molecular biology.

FIG. 8 shows embodiments of a processive enzyme molecular sensor wherein the processive enzyme 84 comprises an exonuclease also having strand displacing activity. As shown, sensor 80 comprises processive enzyme 84 conjugated at one or more attachment points 85 to a bridge molecule 83 bonded between electrode pairs and spanning the electrode gap 89. The bridge molecule 83 comprises first and second ends functionalized to bond to the electrode pair at conjugation points 81 and 82. In this case, the signaling features comprise DNA oligonucleotides 87 hybridized to the template strand 88. The strand displacing exonuclease 84 will displace and remove these oligonucleotides (such as illustrated by the displacement of 87') as it translocates and digests the primary strand 88. Each strand displacement event may generate a signal in the measured current of the circuit, as indicated by the perturbations 801, 802, and 803 in the (i) versus (t) plot. The hybridizing oligonucleotides 87 may comprise DNA or DNA analogs, and may further comprise groups attached to enhance signaling, such as described further below in the context of FIG. 14. In

various embodiments, some exonucleases may displace oligonucleotides 87 from the primary strand 88 by digesting entirely through the double stranded region of the paired segment, or by digesting only partially through it, at which point the remainder of the oligonucleotide dissociates from the primary strand.

5 FIG. 9 shows embodiments of a processive enzyme molecular sensor 90 wherein the processive enzyme 94 comprises a helicase. As shown, sensor 90 comprises processive enzyme 94 conjugated at one or more attachment points 95 to a bridge molecule 93 bonded between electrode pairs and spanning the electrode gap 99. The bridge molecule 93 comprises first and second ends functionalized to bond to the electrode pair at conjugation points 91 and 92. In this case, the signaling features 96, 97 are perturbing groups on the DNA strand 98. As each perturbing group 96, 97 passes by the helicase, it generates a signal in the measured current of the circuit, as indicated by 901, 902 and 903 in the (i) versus (t) plot. As illustrated, the various perturbing groups 96, 97 may be arranged in patterns that encode information. The illustration of this particular sequence of 96, 97, 97, 96, 96, and so
10 forth is not intended to be limiting, and any number of different perturbing groups arranged in any pattern may be used in order to encode information. The DNA molecule may be single stranded, with the perturbing groups 96, 97 on the single strand, and the helicase 94 may be purely translating along the single strand. In other embodiments, the DNA may be double stranded, and the perturbing groups may be positioned on either strand, the one that goes
15 “through” the helicase, to which it is bound, or the one that is displaced “around” the helicase. In other embodiments, such perturbing groups may be positioned on both strands within the same double stranded DNA molecule. Such groups are described further below in the context of FIG. 14.

 FIG. 10 shows embodiments of a processive enzyme molecular sensor 120 wherein
25 the processive enzyme 124 comprises a DNA packaging motor. As shown, sensor 120 comprises processive enzyme 124 conjugated at one or more attachment points 125 to a bridge molecule 123 bonded between electrode pairs and spanning the electrode gap 129. The bridge molecule 123 comprises first and second ends functionalized to bond to the electrode pair at conjugation points 121 and 122. In this case, the signaling features 126, 127 are
30 perturbing groups on the DNA strand 128. As each perturbing group 126, 127 passes by the motor 124, it generates a corresponding signal in a monitored electrical parameter of the circuit, such as indicated by the changes 105, 106, and 107 seen in the (i) versus (t) plot. As illustrated, the various perturbing groups 126, 127 may be arranged in patterns that encode information. The illustration of this particular sequence of 126, 127, 127, 126, 126, and so

forth is not intended to be limiting, and any number of different perturbing groups arranged in any pattern may be used in order to encode information. The DNA molecule may be single stranded, with the perturbing groups on the single strand, such as for a motor that translates a single strand. In other embodiments, the DNA may be double stranded, and the perturbing groups may be positioned on either strand, in the case of a motor that translocates double stranded DNA. Such groups are described further in the context of FIG. 14.

FIG. 11 shows embodiments of processive enzyme molecular sensors 110 wherein the processive enzyme 114 comprises an exonuclease. As shown, sensor 110 comprises processive enzyme 114 conjugated at one or more attachment points 115 to a bridge molecule 113 bonded between electrode pairs and spanning the electrode gap 119. The bridge molecule 113 comprises first and second ends functionalized to bond to the electrode pair at conjugation points 111 and 112. In this case, the signaling features 117 are perturbing groups on the DNA strand 118. As each perturbing group 117 passes near the exonuclease, it generates a signal in a monitored electrical parameter of the circuit, such as indicated by the perturbations 108, 108', and 108'' in the (i) versus (t) plot. The action of the exonuclease 114 on the functionalized DNA 118 produces digestion products 116 as the DNA is digested by the enzyme. The digestion products comprise a mixture of bases and functionalized bases that retain a perturbation group, as shown. The illustration of this particular sequence of perturbing groups 117 is not intended to be limiting, and any number of different perturbing groups arranged in any pattern may be used in order to encode information. The DNA molecule 118 may be single stranded, with the perturbing groups 117 positioned on the single strand, for an exonuclease that digests a single strand, or the DNA may be double stranded, for an exonuclease that processes double stranded DNA, and the perturbing groups may be on either strand, and in the case where only one strand of a double stranded DNA is digested, the perturbing groups 117 may reside on the digested or on the non-digested strand, or both. Such groups are described further below in the context of FIG. 14.

FIG. 12-A shows the detailed protein anatomy and DNA engagement of one of the exemplary processive enzymes for use herein, in this case a polymerase, and specifically the E. Coli Klenow fragment. The structure shown is PDB ID 1KLN. The detailed structure and how it engages the template DNA inform the choice of how to best conjugate the enzyme into the circuit, so as not to interfere with its interaction with DNA, and to position the signaling portions of the protein or DNA near to the molecular bridge for enhanced signal generation via proximity. The structure A at left is the polymerase engaged with a DNA substrate molecule whereas the structure B at right is the polymerase in its conformation absent

engagement with a DNA substrate. The helix, sheet and loop portions of the enzyme are pointed out in both the engaged conformation and the conformation absent DNA engagement.

FIG. 12-B shows the detailed protein anatomy and DNA engagement of one of the exemplary processive enzymes for use herein, in this case a helicase, and specifically the human RECQ-like helicase. The structure shown is PDB ID 2WWY. The detailed structure and how it engages the template DNA inform the choice of how to best conjugate the enzyme into the circuit, so as not to interfere with its interaction with DNA, and to position the signaling portions of the protein or DNA near to the molecular bridge for enhanced signal generation via proximity. The structure shows the helicase engaged with a DNA substrate molecule. The helix, sheet and loop portions of the helicase are pointed out in the illustrated structure.

FIG. 13-A shows embodiments of a processive enzyme molecular sensor 130A wherein the polymerase enzyme 134A is conjugated to a molecular bridge molecule 133A, at a conjugation point 135A comprising a specific site on the enzyme 134A bonded to a specific site on the bridge molecule 133A. As shown, the bridge molecule 133A is bonded to each of the spaced-apart electrodes to span the electrode gap 139A. The bridge molecule 133A comprises first and second ends functionalized to bond to the electrode pair at conjugation points 131A and 132A.

FIG. 13-B shows the molecular structure of one specific example of a processive enzyme molecular sensor 130B, wherein the polymerase 134B is conjugated to a bridge molecule 133B comprising a 20 nm long (= 6 helical turns) double-stranded DNA. The sensor 130B further comprises a pair of spaced apart chromium electrodes 138B and 139B, disposed on a substrate layer, such as SiO₂, and spaced apart at about 10 nm. On each electrode 138B and 139B are deposits of gold 131B that participate in the bonding of the bridge molecule to each of the electrodes. The DNA bridge molecule 133B shown is conjugated to the gold-on-chromium electrodes through thiol groups on first and second ends of the DNA bridge, binding to gold via sulfur-gold bonds 132B, and wherein the polymerase 134B is conjugated to the DNA bridge molecule 133B at a centrally located biotinylated base on the DNA bridge 135B, bound to a streptavidin molecule 136B, in turn bound to the polymerase 134B via a specific biotinylated site 135B on the polymerase 134B. In this way, the streptavidin 136B links the polymerase 134B to the DNA bridge molecule 133B by way of two biotin-streptavidin linkages 135B. The processive enzyme molecular sensor 130B is illustrated translocating a DNA substrate molecule 137B. As discussed, the DNA substrate

molecule 137B may be encoded with information comprising arrangements of signaling features such as bound DNA oligonucleotide segments or perturbing groups.

FIG. 13-C shows embodiments of a processive enzyme molecular sensor 130C wherein the polymerase molecule 134C (also shown in FIG. 12-A) is conjugated directly into the current path between the electrodes, spanning the electrode gap 139C, by conjugation to two separate “arm” molecules 133C and 134C, which in turn are conjugated at one of their ends to each of the two electrodes via attachment points 136C and 137C, and to bonding sites on the enzyme 132C at conjugation points 135C and 138C, as shown. These embodiments illustrate how the conjugations of the arm molecules may be to the ends of an alpha-helix spanning the enzyme, to loop or sheet portions, or to other such protein anatomical structures (e.g., pointed out in FIG 12-A) that preferentially channel the current through or near the active site of the enzyme or through or near portions of the enzyme that experience changes in conformation when interacting with a substrate molecule such as DNA molecule 131C shown.

FIG. 13-D shows embodiments of a processive enzyme molecular sensor 130D wherein the polymerase molecule 132D (also shown in FIG. 12-A) is conjugated directly to each of the electrodes in a pair of spaced-apart electrodes to span the electrode gap 139D without any intervening arm or bridge molecules, in accordance with the general sensor structure of FIG. 4-C. The polymerase 132D is conjugated directly to the electrodes at attachment points 136D and 137D, wherein the attachment points may comprise thiol-gold bonds. These embodiments illustrate how the conjugations may be to the ends of an alpha-helix spanning the enzyme, to loop or sheet portions, or to other such protein anatomical structures (e.g., pointed out in FIG 12-A) that preferentially channel the current through or near the active site of the enzyme or through or near portions of the enzyme that experience changes in conformation when interacting with a substrate molecule such as DNA molecule 131D shown.

FIG. 13-E shows embodiments of a processive enzyme molecular sensor 130E wherein the helicase enzyme molecule 132E (also shown in FIG. 12-B) is conjugated to a molecular bridge molecule 133E, at a specific conjugation point 135E comprising a specific binding site on both the helicase and the bridge molecule, in accordance with the general sensor structure shown in FIG. 4-A. The bridge molecule 133E comprises first and second ends, each bonded to the pair of electrodes at attachment points 136E and 137E. The conjugation point 135E between the helicase 132E and the bridge molecule 133E is chosen to bring the most active or variable parts of the helicase 132E, or the signaling groups bonded on the

DNA substrate molecule 131E to be processively processed by the helicase 132E, in close proximity to the bridge molecule 133E for enhanced current modulation and enhanced signaling.

The electrical parameter measured in the processive enzyme molecular sensors described herein for distinguishable signals can in general be any electrical property of the circuit measurable while the sensor is active. In various embodiments, the parameter is the current passing between the two electrodes in a pair of spaced-apart and bridged electrodes versus time, either continuously or sampled at discrete times, when a voltage, fixed or varying, is applied between the electrodes. In various embodiments, there may also be a gate electrode, capacitively coupled to the molecular structure, such as a buried gate or back gate, which applies a gate voltage, fixed or variable, during the measurement. In various other embodiments, the measured parameter may be the resistance, conductance, or impedance between the two electrodes, measured continuously versus time or sampled periodically. The measured parameter could be the voltage between the electrodes. If there is a gate electrode, the measured parameter could be the gate voltage. The measured parameter could also be a capacitance, or the amount of charge or voltage accumulated on a capacitor coupled to the circuit. The measurement could be a voltage spectroscopy measurement, such that the measurement process comprises capturing an I-V or C-V curve. The measurement could be a frequency response measurement. In all such measurements, for all such measured parameters, there are embodiments in which a gate electrode applies a gate voltage, fixed or variable, near the molecular complex during the measurement. Such a gate will typically be physically located within a micron distance, and in some cases within a 200 nm distance, of the molecular complex bridging the pair of spaced-apart electrodes.

In various embodiments, for the electrical measurements there will be a reference electrode present, such as a Ag/AgCl reference electrode, or a platinum electrode, in the solution placed in contact with the sensor, and maintained at an external potential, such as ground, to maintain the solution at a stable or observed potential, and thereby make the electrical measurements more well defined or controlled. In addition, when making the electrical parameter measurement, various other electrical parameters may be held fixed at prescribed values, or varied in a prescribed pattern, such as, for example, the source-drain electrode voltage, the gate voltage if there is a gate electrode, or the source-drain current.

The use of the present processive enzyme molecular sensors to measure distinguishable features of a DNA molecule requires the processive enzyme to be maintained in the appropriate physical and chemical conditions to be enzymatically active, to process

DNA templates, and produce strong, distinguishable signals above the background noise (i.e., for a high signal-to-noise ratio, or SNR). For this, the enzyme should reside in aqueous buffer solution, which, in various embodiments, will comprise salts, such as NaCl or KCl, pH buffers, such as Tris-HCl, multivalent cation cofactors, such as Mg, Mn, Ca, Co, Zn, Ni, other ions, such as Fe or Cu, surfactants, such as tween, chelating agents such as EDTA, reducing agents such as DTT or TCEP, solvents, such as betaine or DMSO, volume concentrating agents, such as PEG, and/or other components for enzyme buffers. The sensor signals may also be enhanced by maintenance of a buffer in a certain range of pH or temperature, or at a certain ionic strength. In particular, the ionic strength may be selected to obtain a Debye length (electrical charge screening distance) in the solution favorable for electrical signal production, which may be in the range of 0.3 nm—100 nm, or in the range of 1 nm—10 nm. Such buffers formulated to have larger Debye lengths may be more dilute or have lower ionic strength by a factor of 10, 100, 1000, 100,000 or 1 million relative to the buffer concentrations routinely used in standard molecular biology procedures such as PCR.

Buffer compositions, concentrations and conditions (e.g., pH, temperature, or ionic strength) may also be selected or optimized to alter the enzyme kinetics to favorably increase the signal-to-noise ratio (SNR) of the sensor, the overall rate of signal production, or overall rate of information decoding in the context of reading data stored in DNA molecules. This may include slowing down or speeding up the processive enzyme activity by any combination of these variables. Optimal buffer selection process consists of selecting trial conditions from the matrix of all such parameter variations, empirically measuring a figure of merit, such as related to the discrimination of the distinguishable features, or to the overall speed of feature discrimination when processing a template, and using various search strategies, such as those applied in statistical Design Of Experiment (DOE) methods, to infer optimal parameter combinations.

In the case where the processive enzyme is a polymerase, the processing of the template DNA also requires that the polymerase be provided with a supply of dNTPs (deoxynucleoside triphosphates) so that it can act processively on a single-stranded DNA template molecule to synthesize a complementary strand. The standard or native dNTPs are dATP, dCTP, dGTP, and dTTP, which provide the A, C, G, and T base monomers for polymerization into a DNA strand, in the form required for the enzyme to act on them as substrates. Polymerase enzymes, native or mutant, may also accept analogues of these natural dNTPs or modified forms that may enhance or enable the generation of the distinguishable signals in accordance to the present disclosure.

Many such modified forms of dNTPs are known to those skilled in the field of nucleic acid biochemistry, and all such forms may be enabling for signal production in various embodiments of the processive enzyme molecular sensor. This includes dNTPs that have modification to the base, the sugar, and/or the phosphate group. For example, common forms include deaza-, thio-, bromo- and iodo- modifications at various sites on the molecule, or the inclusion of metal ions or different isotopes at various sites, the inclusion of a diverse variety of dye molecules at various sites, or methylation of various sites, or biotinylation of various sites. In particular, such modifications include forms that have an extended phosphate chain beyond the native tri-phosphate, such as to tetra-, penta-, hexa-, hepta- or more (4 or more, up to 11 or more) phosphates. Further, modifications may comprise addition of a chemical group to the terminal phosphate of this chain of phosphates, or any of the phosphates, except the alpha-phosphate or first in the chain, which are cleaved off during incorporation of the dNTPs into a complementary strand.

Polymerases are highly tolerant of such derivatization, and retain a high level of activity in their presence. In various embodiments, such modifying groups may provide different charge states, or different sizes, or different degrees of hydrophobicity to the dNTPs, which may aid in producing distinguishable signals. In various embodiments, groups added to dNTPs may interact selectively with sites on the bridge molecule or on the polymerase or the template DNA to produce distinguishable signals. In various embodiments, the dNTPs may be modified to include a group that interacts directly with the signal generating features indicated in FIG 3-A or FIG-3-B. For example, in the case where oligonucleotides are being displaced from an encoded DNA molecule, the dNTPs may be modified to comprise a group that interacts with the oligonucleotide being displaced, for example via hybridization to a complimentary portion of the oligonucleotide, or via interaction with a cognate group on the oligonucleotide. For example, the binding oligonucleotides and dNTPs may comprise complimentary oligonucleotides, or comprise other cognate binding partners such as biotin-avidin, or pyrenes that engage by pi-stacking, or many other such pairs, with one such partner in the pair being attached to the binding oligonucleotide, and the other cognate partner on a dNTP. In various embodiments, the binding group on the binding oligonucleotide feature of FIG. 3-A is only exposed for binding to its partner when the oligonucleotide is the next strand to be displaced by the polymerase, so that such interactions are preferentially occurring during the signaling process of FIG. 3-A.

Various embodiments of information encoding DNA molecular structures are shown in FIG 14-A. In FIG. 14-A, DNA structure (A) comprises a series of oligonucleotides 142,

144, and so forth, bound to a template strand 140. Such oligonucleotides may be contiguous on the template strand, or have one or more bases such as 141, 143, and so forth separating them. Such oligonucleotides may be native DNA, or may comprise DNA analogues, such as RNA, PNA, LNA, XNA, or modified or analog bases or universal bases such as inosine or 5-nitroindole), or extended genetic code bases such as iso-dC and iso-dG. Such oligonucleotides 142, 144, and so forth may pair in a fully complementary fashion with the template 140, or may include mismatched base pairings. Such oligonucleotides may be selected to have substantially different melting point temperatures (T_m), such as near 30° C, near 40° C, near 50° C, near 60° C or up to about 100° C. The template DNA 140 to which these oligonucleotides bind may comprise native DNA, or comprise analogues such as those just listed.

FIG. 14-A structure (B) is an information encoded DNA molecule in which the signaling features comprise bound oligonucleotides on the primary strand 140B, such as in structure (A), further comprising signal enhancing groups 145, 146, and so forth bound to the oligonucleotides. Such groups 145, 146, and so forth may be conjugated to the oligonucleotides in many ways, including, but not limited to, biotin-avidin binding, click chemistry, or conjugation to a free azide or amine group on the oligonucleotide. Such signal enhancing groups may comprise molecules carrying a formal charge in typical solutions, such as peptides or proteins including streptavidin, or molecules that are sterically bulky or have steric interactions with the processive enzyme, such as a PEG polymer, or molecules that are hydrophobic or hydrophilic, such as various peptides, or molecules that have specific interactions with the processive enzyme, native or genetically modified, or interactions with any of the other components of the molecular complex. Such interactions may be provided by one member of cognate binding partners bonded on the oligonucleotide and the other on the enzyme, the molecular bridge molecule, or the molecular complex. For example, complementary oligonucleotides may be attached to the bound oligonucleotide and the processive enzyme or bridge molecule.

Structures (C) and (D) shown in FIG. 14-A represent further embodiments of DNA molecules encoded with information, such as discussed in the context of FIGS. 3-A and 3-B. In structure (C), signaling features 145', 146', and so forth, such as described for structure (B), are bound to a single stranded DNA template strand 140C. In structure (D), the signaling features 145'', 146'', and so forth are bound to a double-stranded DNA template 140D.

FIG. 14-B illustrates template structure embodiments (or strand architectures) that allow the data payload from a single molecule to be read multiple times by the same

processive enzyme sensor. Various molecular biology methods enable a processive enzyme sensor to interrogate the same DNA molecule repeatedly. Three such embodiments of encoding template structural schematics are shown in FIG. 14-B. The first structural schematic, (A), is a circularized template molecule. For processive enzymes that can engage a circular template without damaging the template, (such as a polymerase, helicase or packaging motors that do not have to “thread” their template but instead clamp on from the side), are able to process around the circular template multiple times, and thus process the same signaling features multiple times. As discussed, such features capable of being processed multiple times include, for example, the permanently bound groups of FIG-14-A (C) and (D), or, in the case of displaceable oligonucleotide features, as in FIG. 14-A (A) and (B), so long as such oligonucleotides are rapidly re-bound to the template after displacement. For oligonucleotides to re-bind to the template, they may be replenished from a local concentration in solution and multiply read. The second structural schematic (B) of FIG 14-B is a hair pinned single strand forming a double stranded region. Suitable processive enzymes that will process one strand of a double stranded molecule, such as strand displacing polymerases, some helicases, some packaging motors, and some exonucleases, go around the hairpin and process the other complementary strand, and thus can in effect read twice the information that is doubly represented as permanently bound groups on both strands. The third structural schematic (C) of FIG 14-B is a DNA encoding molecule wherein the data payload is simply repeated in tandem, one or more times, so that the processive enzyme will trivially process through these multiple copies of the same DNA data payload, thus reading it multiple times.

The DNA molecules comprising encoded information for reading by the cognate processive enzyme molecular sensor, being of synthetic and designed origin, may be prepared with an architecture that facilitates the reading process as well as the encoding (its synthesis) and decoding (the reading) processes. Such architectures are illustrated in FIG. 15 by structural embodiments (A) and (B). Overall, FIG. 15 illustrates structures of information carrying DNA fragments. In structure (A), the START SITE segment is a primer segment containing primer target/structure where the processive enzyme will bind and initiate the processing of the rest of the strand. The L-BUFFER segment may contain signal calibration sequence for the reader, or buffering sequence prior to the DATA PAYLOAD segment, which contains information storing encoded sequence, and related error correction sequence such as parity bits. The R-BUFFER segment may contain additional calibration sequence, as well as buffer sequence allowing for the enzyme to avoid getting too close to the end of the

template when reading data. Left and/or Right buffer regions may comprise DNA segments required to support the polymerase binding footprint, or various calibration or initiation sequences used to help interpret the signals coming from the DATA PAYLOAD region. These buffer segments could also contain molecular barcode sequences that are used to
5 uniquely distinguish molecules, or identity replicate molecules that are derived from the same originating single molecule. One such method of barcoding, known to those skilled in DNA oligonucleotide synthesis, comprises addition of a short random N-mer sequence, typically 1 to 20 bases long, as is commonly made by carrying out synthesis steps with degenerate mixtures of bases instead of specific bases. Finally, there is a DATA PAYLOAD segment,
10 where the specific data is encoded, which may include primary digital data being stored, as well as data related to proper assembly of such information fragments into longer strings, and also data related to error detection and correction, such as parity bits, check sums, or other such information overhead.

In structure (B) of FIG. 15, L-ADAPTER and R-ADAPTER segments may be
15 sequence elements related to the storage or manipulation of the associated DNA segment, such as adapters for outer priming sites for PCR amplification, or hybridization based selection, or representing a surrounding carrier DNA for this insert, including insertion into a host organism genome as a carrier. These right and left adapters may comprise primers for universal amplification processes, used to copy the stored data, or may comprise
20 hybridization capture sites or other selective binding targets, for targeted selection of molecules from a pool, for example. Such adapters may comprise surrounding or carrier DNA, for example in the case of DNA data molecules stored in live host genomes, such as in bacterial plasmids or other genome components of living organisms. The DATA PAYLOAD in general may include the actual primary data being archived, as well as metadata for the
25 storage method, such as related to the assembly of this information into larger strings, or error detection and correction.

The DATA PAYLOAD DNA structure results from a sensor-specific information encoding scheme applied to a source digital data payload, such as binary data, as illustrated in FIG. 16. In this scenario, the originating digital data that is being stored as DNA will
30 typically have a prior representation as electronic binary data. This originating data will then be divided into segments, and be augmented by re-assembly data, and also transformed by error correcting encodings appropriate for DNA data storage, to produce actual binary data payload segments, such as in FIG. 16, that then require translation to DNA payload sequences for subsequent DNA synthesis to produce the physical storage molecules. This

primary translation is what, in various embodiments, is performed by Binary Encoding Schemes (BES) such as those exemplified in FIG. 16. These encoding schemes provide primary translation from a digital data format, such as binary, to a DNA molecular sequence format, via first producing a list of distinguishable signaling features that imply
5 corresponding DNA segments, which are assembled for the encoding DNA molecule.

Choosing which BES is appropriate depends, in part, on the type of distinguishable signal features and their arrangements, as discussed in the context of FIG. 1-B and illustrated in the inset of FIG. 1-B. FIG. 16 illustrates several such primary encodings, beginning with an exemplary binary data payload, a particular 16-bit word “1010100110011100,” and
10 converting the binary data to one or more distinguishable signal features for the encoded DNA molecule.

With continued reference to FIG. 16, exemplary binary encoding schemes (BES) shown include: BES1, encoding 1 bit into 2 distinguishable signaling features F1 and F2, for use with a DNA reading sensor that can distinguish these features; BES2, encoding
15 combinations of two binary bits 00, 01, 10 and 11 into four features, F1, F2, F3 and F4, for use with a DNA reading sensor that distinguishes these features; and BES3, encoding the binary strings 0, 1 and 00 into 3 distinguishable features F1, F2 and F3. FIG. 16 illustrates the encoding of the binary data payload for these BES, in terms of conversion of the binary string into a feature string, which must then be realized in DNA. Codes that produce shorter
20 sequences are preferred if reducing the length of synthesized DNA information encoding molecules is preferred, which may exist for example, due to practical limitations on oligonucleotide length for the writing technology.

It is understood that the BES exemplified in FIG. 16 are non-limiting, and many variations or similar encoding schemes to those shown in FIG. 16 are also implicit in these
25 examples, such as by permuting the features used. It is also understood that all such encoding schemes must have a cognate sensor that is capable of distinguishing the signals of the encoding features, so that the choices of BES are directly related to the properties of the sensor in distinguishing features. It is understood that digital data formats or alphabets other than binary, such as hexadecimal, decimal, ASCII, etc., can equally well be encoded into
30 DNA signaling features by similar schemes as those of FIG. 16. Such methods are well known to those skilled in the field of computer science. Schemes more sophisticated than those shown, in terms of optimal information density, such as Lempel-Ziv encoding, can highly efficiently convert and compress data from one alphabet into another. In general, for converting a binary or other digital data payload string or collection of strings, into a DNA

sequence string, or collection of such strings, many of the methods of lossless and lossy encoding or compression, well known in computer science, can be used to devise schemes for the primary conversion from input digital data payloads to DNA data payloads of the forms such as the embodiments of FIG. 14.

5 For one illustrative example of such how the encoding is used to define a DNA sequence for synthesis, consider a sensor which has as distinguishable signaling features that are the oligonucleotides 5'-CCCC-3' (SEQ ID NO: 3) and 5'-GGGG-3' (SEQ ID NO: 4) and 5'-AAAA-3' (SEQ ID NO: 5) bound to the respective reverse complement template segments F1 = 5'-GGGG-3' (SEQ ID NO: 4), F2 = 5'-CCCC-3' (SEQ ID NO: 3) and F3 = 5'-TTTT-
10 3' (SEQ ID NO: 6). Suppose the BES3 of FIG. 16 is used, and the input binary data payload is 01001. The conversion to a feature sequence would be F1F2F3F1. This could be directly converted to a DNA data payload segment of the DNA encoding molecule with architecture as in FIG. 15, as 5'-GGGGCCCCTTTTGGGG-3' (SEQ ID NO: 7). Or, in other embodiments, there may be "punctuation" sequence segments inserted between the
15 distinguishable signal features, which do not alter the distinguishable features, e.g., bound oligonucleotides, but may provide benefits such as accommodating special properties or constraints of the DNA synthesis chemistry, or to provide spacers for added time separation between signal features, or reduced steric hindrance, or to improve the structure of the DNA molecule. For example, if A were such a punctuation sequence, the DNA encoding sequence
20 would become 5'-AGGGGACCCCATTTTAGGGGA-3' (SEQ ID NO: 8). In general, such insertion of punctuation sequences or filler sequences may be part of the process of translating from a digital data payload to the encoding DNA sequence to be synthesized.

For robust recovery of digital data from DNA storage, a DNA data payload of interest may be processed by a processive enzyme sensor multiple times, or, for a collection of such
25 payloads, they on average may be processed some expected number of multiple times. This repetition has the benefit of providing a more accurate estimation of the encoding distinguishable features by aggregating such multiple observations. This also has the benefit of overcoming fundamental Poisson sampling statistical variability to ensure that, with high confidence, a data payload of interest is sampled and observed at least once, or at least some
30 desirable minimal number of times. In various embodiments, the desired number or expected number of such repeat interrogations may be in the range of 1 to 1000. In other embodiments, the number or expected number may be in the range of 3 to 30.

Such multiple observations may be via either of or a combination of: repeated observations of the same physical DNA molecule by the processive enzyme sensor, or by one

or more processive enzyme sensors processing multiple, physically distinct DNA molecules that carry the same data payload. In the latter case, such multiple, physically distinct DNA molecules with the same data payload may be the DNA molecules produced by the same bulk synthesis reaction, or may be molecules from distinct such synthesis reactions targeting the same data payload, or may be replicate molecules produced by applying amplification or replication methods such as PCR, T7 amplification, rolling circle amplification, or other forms of replication known to those skilled in molecular biology. The aggregation of such multiple observations may be done through many methods, such as averaging or voting, maximum likelihood estimation, Bayesian estimation, hidden Markov methods, graph theoretic or optimization methods, or deep learning neural network methods.

In various embodiments, the digital data stored in DNA is read at a high rate, such as approaching 1 Gigabyte per second for recovering the digital data, as is possible with large scale magnetic tape storage systems. Because the maximum processing speed of a processive enzyme such as a polymerase is in the range of 100-1000 bases per second, depending on the type, the bit recovery rate of one sensor is limited to a comparable speed, and so it is highly desirable to deploy millions of sensors in a cost effective format to achieve the desired data reading capacity. In various embodiments, the sensors can be deployed as a large scale sensor array on a CMOS chip, which is the most cost-effective, semiconductor chip manufacturing process.

FIG. 17-A shows an embodiment of the fabrication stack in which the sensor measurement circuitry is deployed as a scalable pixel array as a CMOS chip, a nano-scale lithography process is used to fabricate the nano-electrodes, and molecular self-assembly chemical reactions, in solution, are used to establish the molecular complex on each nano-electrode in the sensor array. The result of this fabrication stack is the finished DNA reader sensor array chip indicated at the bottom of FIG. 17-A. In certain embodiments of this fabrication stack, the nanoscale lithography is done using a high resolution CMOS node, such as a 28 nm, 22 nm, 20 nm, 16 nm, 14 nm, 10 nm, 7 nm or 5 nm nodes, to leverage the economics of CMOS chip manufacturing. In contrast, the pixel electronics may be done at a coarser node better suited to mixed signal devices, such as 180 nm, 130 nm, 90 nm, 65 nm, 40 nm, 32 nm or 28 nm. Alternatively, the nano-electrodes may be fabricated by any of a variety of other means known to those skilled in the art of nanofabrication, such as e-Beam Lithography or Nano-imprint lithography, or ion beam lithography, or advanced methods of photolithography such any combinations of Extreme UV or Deep UV lithography, multiple patterning or phase shifting masks.

FIG. 17-B illustrates an embodiment of a high-level CMOS chip pixel array architecture for a DNA reader in more detail at the left of the drawing figure. The CMOS chip pixel array architecture comprises a scalable array of sensor pixels, with associated power and control circuitry and major blocks such as Bias, Analog-to-Digital convertors, and timing. The inset in the figure shows an individual sensor pixel as a small bridged structure representing a single processive enzyme molecular sensor, and where this individual electronic sensor is located in the pixel array. FIG. 17-B also illustrates (at the right side of the figure) the details of an embodiment of a processive enzyme molecular electronics sensor circuit pixel in the array. As illustrated, a complete sensor circuit comprises a trans-impedance amplifier, voltage-biasable source, drain, and (optionally) gate electrodes, and a reset switch, along with a processive enzyme electrically connected between the source and drain electrodes (with or without bridge and/or arm molecules). The feedback capacitor illustrated is optional to improve stability of the amplifier. The output of the pixel circuit (the measurable electronic parameter) in this embodiment is current, which is monitored for perturbations relating to the activity of the processive enzyme. That is, the current output from the trans-impedance amplifier is the measurable electrical parameter for this sensor pixel that is monitored for perturbations. It should be noted that one of the two electrodes can be grounded, in which case a biasable voltage is supplied across the electrodes.

FIG. 18 shows an embodiment of a circuit schematic of the pixel amplifier in detail at the left side of the figure, along with simulation results at the right side of the figure showing the voltage signal vs time when used to measure a 10 pA current, and with a reset applied periodically as indicated in the plot. This embodiment exemplifies one non-limiting selection of circuit components and parameters (transistor, resistors, capacitors, etc.).

FIG. 19 illustrates an embodiment of an annotated chip design layout file and the corresponding finished chip for comparison. In FIG. 19, (A), at left, is the finished design of an embodiment of the CMOS pixel array of FIG 17-B with 256 pixels, annotated to show the location of the Bias 190, Array 191 and Decoder 192 regions of the chip. The design layout also comprises a test structures 193 region. In FIG. 19, (B), at right, is a drawing of an optical microscope image of the corresponding finished chip based on the final design, produced at TSMC, Inc. semiconductor foundry (San Jose, CA) with the TSMC 180 nm CMOS process, with no passivation layer.

FIG 20 shows illustrations of scanning electron microscope (SEM) images of the finished CMOS chip 200 of FIG. 19 (256 pixel array, 2 mm x 2 mm), which clearly shows the sub-optical surface features of the 80 μ m pixel 201, and notably the exposed vias (the

source, gate, and drain) where the nano-electrodes can be deposited by post-processing and electrically connected into the amplifier circuit as shown in FIG 17-B, at right. The furthest right drawing of a 100 nm SEM image 202 in Figure 20 shows an e-beam lithography fabricated pair of spaced apart nanoelectrodes with a molecular complex in place. The sketch 5 203 at the bottom right of Figure 20 is an illustration of the processive enzyme molecular electronics sensor comprising a polymerase molecular complex 207, spaced apart electrodes 204 and 205, each labeled by a gold dot contact, wherein the electrode gap 206 is about 10 nm.

In various embodiments of the present disclosure, a DNA reader chip for use herein 10 comprises at least 1 million sensors, at least 10 million sensors, at least 100 million sensors, or at least 1 billion sensors. Recognizing that a typical sensor data sampling rate may be 10 kHz, and recording 1 byte per measurement, a 100 million sensor chip produces raw signal data at a rate of 1 Terabyte (TB) per second. In considering how many sensors are desirable on a single chip, one critical consideration is the rate at which such a chip can decode digital 15 data stored in DNA, compared to the desirable digital data reading rates. It is, for example, desirable to have digital data read out at a rate of up to 1 Gigabyte per second. Note that each bit of digital data encoded as DNA will require multiple signal measurements to recover, given that a feature of the signal use used to store this information, so this raw signal data production rate for the measured *signal* will be much higher than the recovery rate of encoded 20 digital data. For example, if 10 signal measurements are required to recover 1 bit of stored digital data, as might be the case for signal features such as in FIG. 2, and each measurement is an 8-bit byte, that is a factor of 80 bits of signal data to recover 1 bit of stored digital data. Thus, digital data reading rates are anticipated to be on the order of 100 times slower than the sensor raw signal data acquisition rate. For this reason, achieving desirable digital data 25 reading rate of 1 Gigabyte / second would require nearly 0.1 TB / second of usable raw signal data. Given that not all sensors in a chip may be producing usable data, the need for chips that produce up to 1 TB/sec of raw data is desirable, based on the desired ultimate digital data recover rates from data stored as DNA. In various embodiments, such recovery rates correspond to a 100 million sensor chip.

30 In various embodiments of the present disclosure, multiple chips are deployed within a reader system to achieve desired system-level digital data reading rates. The DNA data reader chip of FIG. 17-A may be deployed as part of a complete system for reading digital data stored in DNA. The features of such a system are indicated in FIG. 22. In various aspects, and with reference to FIG. 22, a complete digital data reading system comprises a

motherboard with a staging area for an array of multiple chips, in order to provide data reading throughput beyond that of the limitations of a single chip. Such chips are individually housed in flow cells, with a fluidics liquid handling system that controls the additional and removal of the sensor system liquid reagents. In addition, the fluidics system receives DNA
5 encoding data in solution form, originating from a data repository source. The motherboard would also comprise a suitable first stage data processing unit, capable of receiving and reducing raw signal data at very high rates, such as exceeding 1TB per second, or exceeding 10 TB per second, or exceeding 100 TB per second, indicated as a primary signal processor. This primary processor may comprise one, multiple, or combinations of a FPGA, GPU, or
10 DSP device, or a custom signal processing chip, and this may optionally be followed by stages of similar such signal processors, for a processing pipeline. Data output of this primary pipeline is typically transferred to a fast data storage buffer, such as a solid state drive, with data from here undergoing further processing or decoding in a CPU-based sub-system, from which data is buffered into a lower speed mass storage buffer, such as a hard
15 drive or solid state drive or array of such drives. From there it is transferred to an auxiliary data transfer computer sub-system that handles the subsequent transfer of decoded data to a destination. All these system operations are under the high-level control of an auxiliary control computer that monitors, coordinates and controls the interplay of these functional units and processes.

20 In some embodiments, chips within the reader system may be disposable, and replaced after a certain duty cycle, such as 24 hours to 48 hours. In other embodiments, the chips may be reconditioned in place after such a usage period, whereby the molecular complex, and possibly conjugating groups, are removed, and then replaced with new such components through a series of chemical solution exposures. The removal process may
25 comprise using voltages applied to the electrodes to drive removal, such as an elevated voltage applied to the electrodes, or an alternating voltage applied to the electrodes, or a voltage sweep. The process may also comprise the use of chemicals that denature, dissolve or dissociate or otherwise eliminate such groups, such as high molarity Urea, or Guanidine or other chaotropic salts, proteases such as Proteinase K, acids such as HCl, bases such as KOH
30 or NaOH, or other agents well known in molecular biology and biochemistry for such purposes. This process may also include the use of applied temperature or light to drive the removal, such as elevated temperature or light in conjunction with photo-cleavable groups in the molecular complex or conjugation groups.

FIG. 22 illustrates an embodiment of a cloud based DNA data archival storage system, in which the complete reader system such as outlined in FIG. 21 is deployed in aggregated format to provide the cloud DNA reader server of the overall archival storage and retrieval system. FIG. 22 shows a cloud computing system, with a standard storage format (upper left).
5 Such a standard cloud computing system is provided with DNA archival data storage capability as indicated. Some cloud-based DNA synthesis system can accept binary data from the cloud computer, and produce the physical data encoding DNA molecules. This server stored the output molecules in a DNA data storage archive, lower right, where typically the physical DNA molecules that encode data could be stored in dried or lyophilized format, or in
10 solution, at ambient temperature or cooled or frozen. From this archive, when data is to be retrieved, a DNA sample from the archive is provided to the DNA data reader server, which outputs decoded binary data back to the primary cloud computer system. This DNA data reader server may be powered by a multiplicity of DNA reader chip-based systems of the kind indicated in FIG. 21, in combination with additional computers that perform the final
15 decoding of the DNA derived data back to the original data format of the primary cloud storage system.

Alternative embodiments of processive enzyme molecular sensors relying on a different nano-electronic measurement configuration than the sensors of FIG 1-B, is shown in FIG. 23. As shown, the processive enzyme molecular sensor 230 comprises a membrane 231
20 disposed between positive and negative electrodes. In these sensors, the electronic measurement is made by a nanopore ionic current sensor comprising the electrodes on either side of the membrane 231, a pore 232 localized in the membrane 231 that regulates the passage of ionic current (shown as "CURRENT (i)" passing through the pore), and aqueous solution residing on both sides of the pore 232. The sensor 230 further comprises a processive
25 enzyme 234 bonded to another molecule 235 embedded in the membrane 231, or bonded directly to the membrane 231. As the processive enzyme 234 processes the DNA substrate 233 having distinguishable signaling features bonded thereto, ions generated on the processing side of the membrane 231 pass through the pore 232 due to the ion gradient. Such nanopore current sensors are well known to those skilled in the art of biological ion channels,
30 or biophysics. In various embodiments, the pore 232 comprises a biological protein nanopore, native or mutated, and the membrane 231 comprises a lipid membrane, or synthetic analogue thereof. Various pores may comprise a solid state pore, and the membrane a thinned membrane composed of a solid material, such as SiN or Teflon. In various embodiments, the positive and negative electrodes may be reversed from what is illustrated, such as depending

on the formal charge of the ions in the ion gradient. As the ions generated in the processive action of the enzyme pass through the pore, the changes in the ionic current are seen as perturbations 236, 237, 238, and so forth in the (i) versus (t) plot shown in the inset in the figure. As shown in FIG. 23, the processive enzyme 234 is further complexed with the pore 232 by a variety of ways, such as being part of a molecular complex involving a small number of molecules 235 embedded in the membrane 231 and associated with the pore 232. As the enzyme 234 processes a DNA template 233, with signaling groups or features shown bonded thereon, the ionic current through the pore is modulated by this activity, producing distinguishable signal perturbations 236, 237, 238, and so forth that correspond to the distinct template signaling features on the DNA template.

Other embodiments of the nanopore current sensor version of the processive enzyme-based DNA digital data reader are shown in FIG. 24. In various embodiments, the processive enzyme molecular sensor 240 comprises a processive enzyme 244 directly conjugated to the pore 242 in the membrane 241 at a specific conjugation point 245. The processive enzyme 244 may comprise a helicase. When a DNA template 243 encoded with signaling features 246 is provided to the enzyme 244, distinguishable signals, (248, 248', 248'', and so forth), result in the (i) versus (t) plot as shown in the inset in the drawing figure, comparable to the situation illustrated in FIG. 6. In certain aspects, the oligonucleotides 246 bonded to the primary strand 243 may comprise attached groups 247 capable of occluding the pore 242 while the oligonucleotides 246 engage with the enzyme 244, thereby resulting in current suppression features. As illustrated, group 247' is capable of occluding the pore 242 when adjacent to, or processed by, the processive enzyme 244. Once processed, an oligonucleotide with its attached group, such as 246' shown, is displaced from the primary strand 143. The enzyme conjugation 245 to the pore 242 may comprise any of many possible conjugation chemistries, such as the Spy-SpyCatcher protein-based conjugation system. In other examples, the conjugation may comprise a molecular tether. For the nanopore sensor embodiments illustrated in FIG. 23 and FIG. 24, various aspects put forth above in the context of FIG. 1-B also apply, providing a nanopore ion current sensor-based sensor for reading digital data stored in DNA molecules, and the related beneficial aspects, encoding schemes, chip formats, systems and cloud based DNA digital data storage systems.

In various embodiments of a processive enzyme molecular sensor, the sensor comprises a carbon nanotube as the bridge molecule, as illustrated in FIG. 25. As shown, the processive enzyme molecular sensor 250 comprises a carbon nanotube 251 connected at each of its ends to spaced apart electrodes to span the electrode gap 256. In certain aspects, the

carbon nanotube 251 comprises a single or multi-walled nanotube, conjugated to the processive enzyme 252 at a specific conjugation site 255, using any of many possible conjugation chemistries. Such a conjugation 255 may, for example, comprise a pyrene linker attached to the nanotube via pi-stacking of the pyrene on the nanotube, or may comprise an attachment to a defect site residing in the carbon nanotube. As per previous embodiments, the processive enzyme 252 processes a DNA template comprising a primary strand 253 with bonded oligonucleotides 254 functioning as distinguishable signaling features that result in perturbations 257, 258, 259, and so forth in the (i) versus (t) plot as shown in the inset of the figure. As the processive enzyme 252 processes the DNA template, oligonucleotides such as 254' shown are displaced. In these embodiments, the current passing through the carbon nanotube 251, acting as a molecular wire, is known to be highly sensitive to the molecules in the environment around the nanotube (e.g. as indicated in FIG. 1-B). It is further known to be sensitive to the activity of an enzyme molecule properly conjugated to the nanotube, including polymerase enzymes. For the embodiments of the sensor 250, all the aspects of the invention put forth above apply in this instance, to provide a carbon nanotube based sensor for reading digital data stored in DNA molecules, including the related beneficial aspects, encoding schemes, chip formats, systems and cloud based DNA digital data storage systems.

In various embodiments, a processive enzyme molecular sensor provides optical signals via an internal Zero Mode Waveguide sensor, as shown in FIG. 26. Such a sensor 260 comprises a single processive enzyme 262 conjugated at conjugation site 265 to the bottom of a metallic well 268, in the evanescent zone of the excitation field applied to the thin substrate, in a Total Internal Reflection mode. The processive enzyme 262 is provided with an encoded DNA template having a primary strand 263 with dye labels 266 on the bound oligonucleotides 264 on the template. When such a dye label 266 is translocated to the enzyme 262, the dye label 266 is held in the evanescent field, and is stimulated by the excitation beam 261 to emit photons 267 of the corresponding dye energy spectrum or color. The result is that, under appropriate conditions, such a sensor produces distinguishable optical signals 269, 269', 269'' and so forth, as indicated in a percent (%) transmission (T) versus time (t) plot, which can be used to encode digital information into DNA molecules. The distinguishable signals here may be photon emissions of a different energy distribution, or color, or emissions with different distinguishable spectra, or different duration or intensity or shape of the spectra versus time, or any combination of such elements that result in distinguishable features. For the Zero Mode Waveguide sensor embodiments indicated in FIG. 26, all the aspects of the invention put forth above in the context of FIG. 1-B also apply

in this instance, to provide a Zero Mode Waveguide-based sensor for reading digital data stored in DNA molecules, and the related beneficial aspects, encoding schemes, chip formats (in this case, optical sensor chips, such as image sensor chips), systems and cloud based DNA digital data storage systems may apply to such a sensor.

5 In other embodiments, a molecular sensor used for reading encoded information may comprise an enzyme molecule capable of processing a non-DNA polymer template, as long as such a polymer allowed for bound, displaceable groups, or for permanently bound signaling features, as discussed above in the context of DNA molecules with oligonucleotides or perturbation groups bonded to a primary strand. Various embodiments of non-DNA data
10 reading sensors is shown in FIG. 27, wherein a lysozyme enzyme 272 is attached to a molecular bridge 274 at specific conjugation point 275. The bridge molecule comprises first and second ends for bonding to each electrode in a pair of spaced apart electrodes at contact points 276 and 277. In these sensors 270, the lysozyme 272 sequentially digests a peptidoglycan polymer 271 having amino acid side chains into digestion products 278
15 comprising various saccharides. Digital data may be encoded into the peptidoglycan 217, by bound groups 273, which are read in a fashion similar to the sensor of FIG 11. Thus, polymers other than DNA can be used to store digital data, with readers as outlined herein, and with a cognate processive enzyme that translocates along such a polymer as all or part of its enzyme activity.

20 Further Embodiments

 In various embodiments, a DNA data reading sensor, usable in a DNA data storage system, is disclosed. The molecular sensor comprises: (a) a pair of spaced-apart nano-electrodes; (b) a single processive enzyme complexed between the nano-electrodes to form a molecular electronic circuit; (c) a measurable electronic parameter of the circuit, modulated
25 by the enzyme activity; (d) a measurement buffer and reagent solution and electrical operating parameters for making such measurements of the parameter of (c); and a feature of a DNA template molecule to be processed by the enzyme, with at least two possible states, that, when processed by enzyme of (b), said states produce distinguishable electrical signals in the measurable parameter of (c), when performed in the conditions provided by buffer and
30 settings of (d).

 The DNA data reading sensor may further comprise a gate electrode.

 The processive enzyme of the DNA data reading sensor may be conjugated in place using a bridge molecule, or any number of arm or linker molecules to assist bonding the enzyme into the sensor circuit, such as shown in the drawing figures, and in particular where

such bridge or arm molecules comprise a double stranded DNA, a protein alpha helix, a graphene nanoribbon, a carbon nanotube, an antibody, or Fab arm of an antibody.

The processive enzyme of the DNA data reading sensor may comprise a native or genetically engineered form of one of: a polymerase, a reverse transcriptase, a helicase, an exonuclease, and a molecular motor for packaging of viral DNA.

The measurable electronic parameter of the DNA data reading sensor is the source-drain current between the electrodes.

The signaling features may comprise oligonucleotides bound to a DNA template, which are displaced by the processive enzyme as the enzyme translocates along the DNA.

The signaling features can also comprise chemical groups strongly bound to the DNA template such that they are not displaced from the DNA template strand as the enzyme translocates along the DNA.

The signaling features can also comprise oligonucleotides bound to the DNA template, wherein the oligonucleotides each have an additional chemical group attached, and wherein the oligonucleotides are displaced by the processive enzyme as it translocates along the DNA.

The DNA segments that produce distinguishable electrical signals may comprise any combination of binding sites for oligonucleotides and chemical groups conjugated to the DNA.

In various embodiments, a CMOS sensor array chip is disclosed. The CMOS sensor array chip may comprise an array of the DNA data reading sensor described along with supporting pixel circuitry that performs measurement of the measurable electrical parameter.

In various embodiments, a method of reading data encoded in a DNA molecule is disclosed. The method comprises: obtaining a DNA molecule that uses distinguishable signals of the DNA data reading sensor described to encode digital data; applying the molecule to the sensor within the buffer and reagent solution described; measuring the measurable electronic parameter; recording or extracting or capturing the distinguishable signals; and converting the measured distinguishable signals into the encoded data format.

In various embodiments, a method of reading data encoded in DNA with enhanced accuracy is disclosed. The method comprises: obtaining a DNA molecule that uses the distinguishable signals of the DNA data reading sensor described above to encode digital data; applying the molecule to the sensor of claim within the buffer and reagent solution described; measuring the measurable electronic parameter as described; repeating these steps multiple times for the same data payload of the DNA, through any combination of the

following: distinct sensors capturing and processing distinct DNA molecules containing the same data payload; the same sensor processing the same DNA molecule multiple times, by using such molecules with a circular or hairpin or tandem repeat architecture that enables multiple processing; and the same sensor capturing and processing distinct DNA molecules that contain the same data payload; recording or extracting or capturing the distinguishable signals from all such reads of the same data payload; and converting the measured distinguishable signals from the multiple reads for the data payload into the encoded data format, by algorithmically aggregating the multiple readings to produce a most accurate, inferred encoded data format of the data payload.

5
10 A method of encoding data into a DNA molecule uses the distinguishable signal features of DNA from a DNA data reading sensor.

A method of encoding data into a DNA molecule uses a series of hybridized oligonucleotides, wherein such oligonucleotides may be with or without added chemical groups.

15 A method of encoding data into a DNA molecule uses encoding schemes in accordance with those described in FIG. 16 and related text, wherein said features produce distinguishable signal features in a DNA data reading sensor as described, and wherein such encoding comprises use of a bound oligonucleotide and/or use of a conjugated chemical group.

20 A method of encoding data into a DNA molecule using distinguishable signal features of a DNA data reading sensor comprises: treating distinguishable features as information states or symbols of an alphabet; using a lossless or lossy data encoding scheme that translates data encoded as a string in a binary or other digital alphabet, to a string in this feature symbol representation; and defining the encoded DNA sequence in accordance with the resulting string of features, by directly transforming it into the corresponding sequence, or possibly with standard other sequence elements placed between such features, such as spacers, or elements related to the method of DNA synthesis, or other forms of punctuation sequence between distinguishable signal feature sequences.

25
30 A method for reading data encoded in DNA in parallel comprises: applying a multiplicity of DNA molecules that use the distinguishable signal DNA features of a DNA data reading sensor as described to encode digital data; applying said molecules to a sensor array chip as described within a buffer and reagent solution; measuring the measurable electronic parameter of for each sensor within the array using the pixel circuitry mentioned; recording or extracting or capturing the distinguishable signals from each pixel for each DNA

molecule; and converting the measured distinguishable signals from each pixel into the encoded data format for each DNA molecule.

5 A method for reading data stored in DNA at very high throughput comprises: supplying one or more chips to a motherboard that controls and transfers data from said chips; introducing DNA molecules encoding information utilizing the distinguishable signal features of the chip, and the related measurement buffer, to the sequencing chips; and capturing the resulting distinguishable signal features for the molecules of (b) and converting them back to the digital data encoded format.

10 A DNA data reading system comprises one or more such reading sensor array chips as described; an electronic hardware system for controlling and managing the electrical inputs and data outputs of such chips; a fluidic system for introducing the data encoding DNA molecules and measurement buffer to the chips; and a signal processing and data recording system for capturing the measured distinguishable signals, and converting these distinguishable signal feature measurements back to the digital data encoded format.

15 The chips may be single use, and replacement chips can be loaded to provide reading capacity over time.

The chips can be reconditioned or reset *in situ*, for repeated or continuous use to provide reading capacity over time.

20 In various embodiments, a cloud based DNA data storage system is disclosed. The system comprises: a data reader server cloud, which in turn comprises a DNA data reading system as described.

25 In various embodiments, a molecular electronics DNA data reading sensor is disclosed. The sensor comprises: a nanopore ion current sensor; a single processive enzyme complexed with the nanopore as part of a molecular complex; a measurement buffer and reagent solution and electrical operating parameters for making such measurements of the ion current through the pore; and a feature of the DNA template molecule, with at least two possible states, that, when processed by enzyme of (b), said states produced distinguishable electrical signals in the nanopore ion current sensor of (a), when performed in the conditions provided by buffer and settings of (c).

30 The signaling features may comprise oligonucleotides with attached groups that alter power ion current by directly engaging with the pore.

These sensors described can be used to read data encoded in DNA, in accordance with the methods disclosed.

In various embodiments, a molecular electronics DNA data reading sensor is disclosed. The sensor comprises: a pair of nano-electrodes; a carbon nanotube spanning the electrodes and in electrical contact; a single processive enzyme conjugated to the carbon nanotube; and a measurement buffer and reagent solution and electrical operating parameters
5 for making measurements of the current passing through the nanotube between the electrodes.

In various embodiments, a DNA data reading sensor comprises: (a) a Zero Mode Waveguide; (b) a single processive enzyme conjugated to the bottom of the zero mode waveguide; (c) a measurement buffer and reagent solution and optical excitation operating
10 parameters for making a measurements of optical emissions from the Zero Mode Waveguide; and (d) a feature of the DNA template molecule, with at least two possible states, that, when processed by enzyme of (b), said states produced distinguishable optical signals when performed in the conditions provided by buffer and settings of (c).

The above sensor may be used to read data encoded in DNA, in accordance with the
15 methods described.

In various embodiments, a molecular electronics sensor for reading digital data encoded in a polymer is disclosed. The sensor comprises: a pair of nano-electrodes; a non-DNA polymer; a single processive enzyme complexed between the nano-electrodes as part of a molecular electronic circuit, where said enzyme acts processively on the polymer; a
20 measurable electronic parameter of the circuit, modulated by the enzyme activity; a measurement buffer and reagent solution and electrical operating parameters for making such measurements of the measurable electronic parameter; and a feature of the polymer molecule, with at least two possible states, that, when processed by processive enzyme, said states produced distinguishable electrical signals in the measurable electronic parameter when
25 performed in the conditions provided by the buffer and settings.

This sensor may be used to read data encoded in the polymer in accordance with the methods described, wherein DNA in the prior embodiments is replaced by the polymer.

DNA data reading processive enzyme molecular sensors and methods of making and using same are provided. References to "various embodiments", "one embodiment", "an
30 embodiment", "an example embodiment", etc., indicate that the embodiment described may include a particular feature, structure, or characteristic, but every embodiment may not necessarily include the particular feature, structure, or characteristic. Moreover, such phrases are not necessarily referring to the same embodiment. Further, when a particular feature, structure, or characteristic is described in connection with an embodiment, it is submitted that

it is within the knowledge of one skilled in the art to affect such feature, structure, or characteristic in connection with other embodiments whether or not explicitly described. After reading the description, it will be apparent to one skilled in the relevant art(s) how to implement the disclosure in alternative embodiments.

5 Benefits, other advantages, and solutions to problems have been described with regard to specific embodiments. However, the benefits, advantages, solutions to problems, and any elements that may cause any benefit, advantage, or solution to occur or become more pronounced are not to be construed as critical, required, or essential features or elements of the disclosure. The scope of the disclosure is accordingly to be limited by nothing other than
10 the appended claims, in which reference to an element in the singular is not intended to mean "one and only one" unless explicitly so stated, but rather "one or more." Moreover, where a phrase similar to 'at least one of A, B, and C' or 'at least one of A, B, or C' is used in the claims or specification, it is intended that the phrase be interpreted to mean that A alone may be present in an embodiment, B alone may be present in an embodiment, C alone may be
15 present in an embodiment, or that any combination of the elements A, B and C may be present in a single embodiment; for example, A and B, A and C, B and C, or A and B and C.

All structural, chemical, and functional equivalents to the elements of the above-described various embodiments that are known to those of ordinary skill in the art are expressly incorporated herein by reference and are intended to be encompassed by the present
20 claims. Moreover, it is not necessary for a device or method to address each and every problem sought to be solved by the present disclosure, for it to be encompassed by the present claims. Furthermore, no element, component, or method step in the present disclosure is intended to be dedicated to the public regardless of whether the element, component, or method step is explicitly recited in the claims. No claim element is intended
25 to invoke 35 U.S.C. 112(f) unless the element is expressly recited using the phrase "means for." As used herein, the terms "comprises", "comprising", or any other variation thereof, are intended to cover a non-exclusive inclusion, such that a molecule, composition, process, method, or device that comprises a list of elements does not include only those elements but may include other elements not expressly listed or inherent to such molecules, compositions,
30 processes, methods, or devices.

We Claim:

1. A sensor comprising:
 - a first electrode;
 - a second electrode spaced apart from the first electrode by an electrode gap;
 - 5 a processive enzyme conjugated to the first and second electrodes, the processive enzyme comprising a native or genetically engineered polymerase, reverse transcriptase, helicase, exonuclease, or molecular motor for packaging of viral DNA; and
 - a trans-impedance amplifier electrically connected to at least one of the first electrode and second electrode, the trans-impedance amplifier providing an output comprising a
 - 10 measurable electrical parameter;
 - wherein the measurable electrical parameter comprises distinguishable signals corresponding to enzymatic activity of the processive enzyme.

2. The sensor of claim 1, wherein the trans-impedance amplifier provides a biasable voltage
- 15 across the first and second electrodes, and wherein the measurable electrical parameter comprises a current output.

3. The sensor of claim 1, further comprising a gate electrode capacitively coupled to the
- 20 electrode gap.

4. The sensor of claim 3, wherein the trans-impedance amplifier further provides a biasable voltage to the gate electrode.

5. The sensor of claim 1, wherein the processive enzyme is directly wired between first and
- 25 second electrodes to provide a conductive pathway between first and second electrodes, through the processive enzyme.

6. The sensor of claim 1, wherein the processive enzyme is conjugated to the first and second
- 30 electrodes via a bridge molecule spanning the electrode gap, the bridge molecule having first and second ends, the first end conjugated to the first electrode and the second end conjugated to the second electrode, wherein the processive enzyme is conjugated to the bridge molecule.

7. The sensor of claim 6, wherein the bridge molecule comprises a double stranded DNA, a protein alpha helix, a graphene nanoribbon, a carbon nanotube, an antibody, or a Fab arm of an antibody.
- 5 8. The sensor of claim 1, wherein the processive enzyme is conjugated to the first and second electrodes via at least one intervening arm molecule, the arm molecule conjugated to at least one of the first and second electrodes and conjugated to the processive enzyme.
9. The sensor of claim 8, wherein the arm molecule comprises a double stranded DNA, a protein alpha helix, a graphene nanoribbon, a carbon nanotube, an antibody, or a Fab arm of an antibody.
- 10 10. The sensor of claim 1, wherein first and second electrodes comprise source and drain electrodes, respectively, and wherein the measurable electrical parameter is the source-drain current between the electrodes.
- 15 11. The sensor of claim 10, wherein the trans-impedance amplifier further provides for voltage-biasable source, drain and gate electrodes.
- 20 12. A method of reading encoded information, the method comprising:
translocating a synthetic DNA molecule through a processive enzyme, the processive enzyme electrically connected in a molecular sensor circuit; and
generating signals in a measureable electrical parameter of the circuit,
wherein the signals correspond to the encoded information.
- 25 13. The method of claim 12, wherein the processive enzyme comprises a native or genetically engineered polymerase, reverse transcriptase, helicase, exonuclease, or molecular motor for packaging of viral DNA.
- 30 14. The method of claim 12, wherein the encoded information comprises binary data.
15. The method of claim 12, wherein the synthetic DNA molecule comprises a DNA template strand with distinguishable signaling features bonded thereon.

16. The method of claim 15, wherein the distinguishable signaling features comprise a sequence of oligonucleotides.

17. The method of claim 16, wherein the oligonucleotides are bonded to the DNA template strand by complementary base pairing, the oligonucleotides are displaced from the DNA
5 template strand by the processive enzyme as the synthetic DNA molecule processively translocates through the processive enzyme, and wherein a signal in the measurable electrical parameter of the circuit corresponds to displacement of an oligonucleotide from the DNA template strand.

10

18. The method of claim 17, wherein each of the oligonucleotides further comprises at least one chemical group bonded thereon.

19. The method of claim 16, wherein the oligonucleotides are covalently bonded to the DNA
15 template strand, and wherein a distinguishable signal in the measurable electrical parameter of the circuit corresponds to interaction of an oligonucleotide with the processive enzyme without displacement of the oligonucleotide from the DNA template strand.

20. The method of claim 19, wherein each of the oligonucleotides further comprises at least
20 one chemical group bonded thereon.

21. The method of claim 15, wherein the distinguishable signaling features comprise chemical groups conjugated to the DNA template strand.

25 22. A DNA information system comprising:

a synthetic DNA molecule comprising a DNA template strand and distinguishable signaling features bonded thereon, the distinguishable signaling features encoding information in the synthetic DNA molecule;

a buffer solution in contact with the synthetic DNA molecule, the first and second
30 electrodes, and the processive enzyme; and

a sensor capable of reading the information encoded in the synthetic DNA molecule, the sensor comprising: a first electrode; a second electrode spaced apart from the first electrode by an electrode gap; a processive enzyme conjugated to both the first and second electrodes; and a trans-impedance amplifier electrically connected to at least one of the first

and second electrodes, the trans-impedance amplifier providing an output comprising a measurable electrical parameter,

wherein the measurable electrical parameter comprises distinguishable signals corresponding to the encoded information.

5

23. The system of claim 22, wherein the processive enzyme comprises a native or genetically engineered polymerase, reverse transcriptase, helicase, exonuclease, or molecular motor for packaging of viral DNA.

10 24. The system of claim 22, wherein the encoded information comprises binary data.

25. The system of claim 22, wherein the distinguishable signaling features comprise a sequence of oligonucleotides.

15 26. The system of claim 25, wherein the oligonucleotides are bonded to the DNA template strand by complementary base pairing.

27. The system of claim 26, wherein each of the oligonucleotides further comprises at least one chemical group bonded thereon.

20

28. The system of claim 25, wherein the oligonucleotides are covalently bonded to the DNA template strand.

25 29. The system of claim 28, wherein each of the oligonucleotides further comprises at least one chemical group bonded thereon.

30. The system of claim 22, wherein the distinguishable signaling features comprise chemical groups conjugated to the DNA template strand.

30 31. The system of claim 22, further comprising a reference electrode submerged in the buffer solution.

32. The system of claim 22, further comprising a gate electrode capacitively coupled to the electrode gap.

33. The system of claim 22, wherein the processive enzyme is directly wired between first and second electrodes to provide a conductive pathway between first and second electrodes, through the processive enzyme.

5

34. The system of claim 22, wherein the processive enzyme is conjugated to the first and second electrodes via a bridge molecule spanning the electrode gap, the bridge molecule having first and second ends, the first end conjugated to the first electrode and the second end conjugated to the second electrode, wherein the processive enzyme is conjugated to the
10 bridge molecule.

35. The system of claim 34, wherein the bridge molecule comprises a double stranded DNA, a protein alpha helix, a graphene nanoribbon, a carbon nanotube, an antibody, or a Fab arm of an antibody.

15

36. The system of claim 22, wherein the processive enzyme is conjugated to the first and second electrodes via at least one intervening arm molecule, the arm molecule conjugated to at least one of the first and second electrodes and conjugated to the processive enzyme.

20

37. The sensor of claim 36, wherein the arm molecule comprises a double stranded DNA, a protein alpha helix, a graphene nanoribbon, a carbon nanotube, an antibody, or a Fab arm of an antibody.

38. The system of claim 22, further comprising a CMOS sensor array chip comprising an
25 array of the sensors and supporting pixel circuitry that performs measurement of the measurable electrical parameter.

39. The system of claim 38, comprising at least two of the CMOS sensor array chips; an electronic hardware system for controlling and managing electrical inputs and data outputs of the chips; a fluidic system for introducing the synthetic DNA molecule in the buffer solution
30 to the chips; and a signal processing and data recording system for capturing the distinguishable signals and for converting the distinguishable signals back to the information.

40. The system of claim 22, wherein the synthetic DNA molecule comprises a circular, hairpin or tandem repeat architecture that allows repeat reading of the information encoded in the synthetic DNA molecule.

5

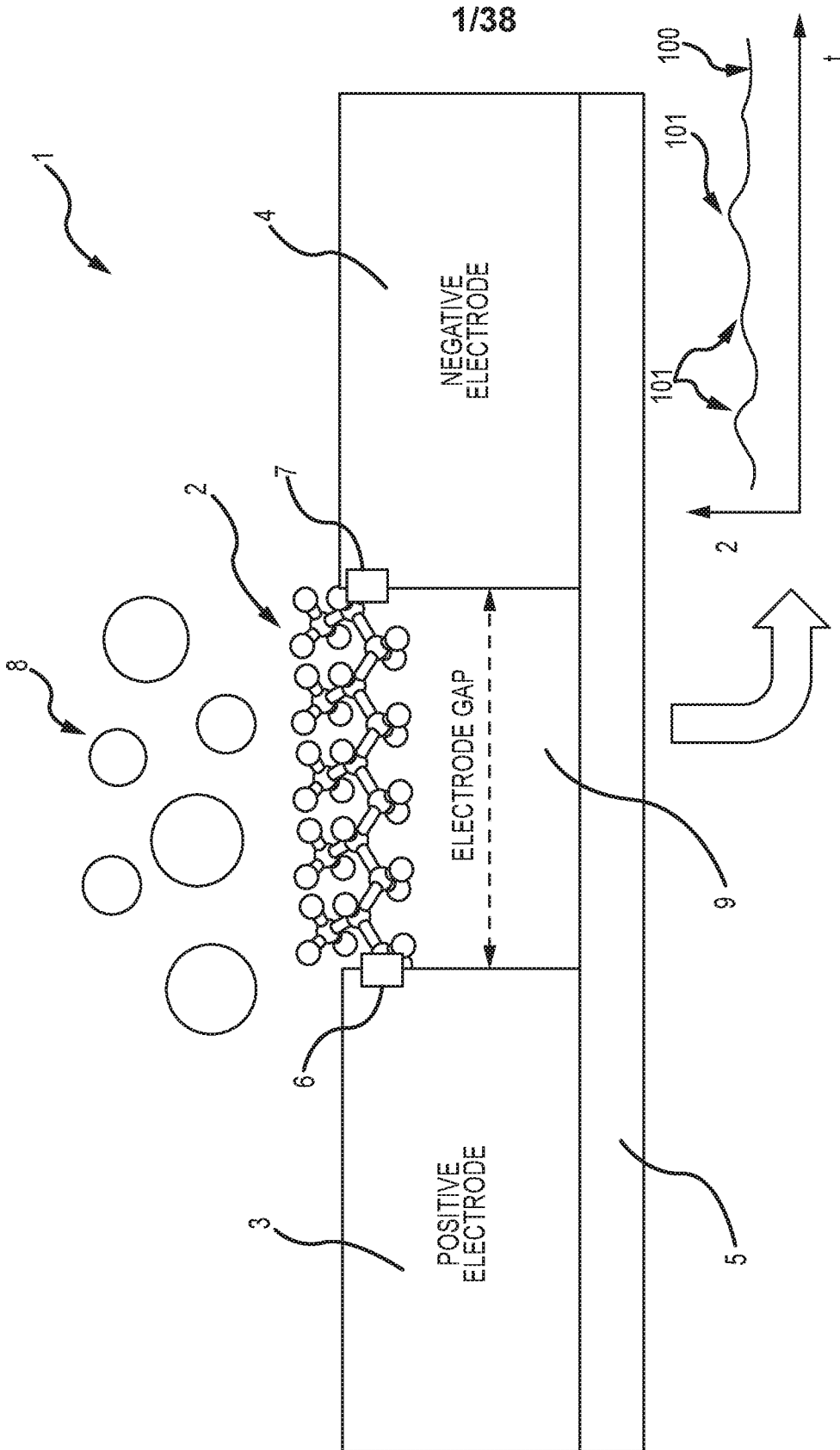


FIG.1-A

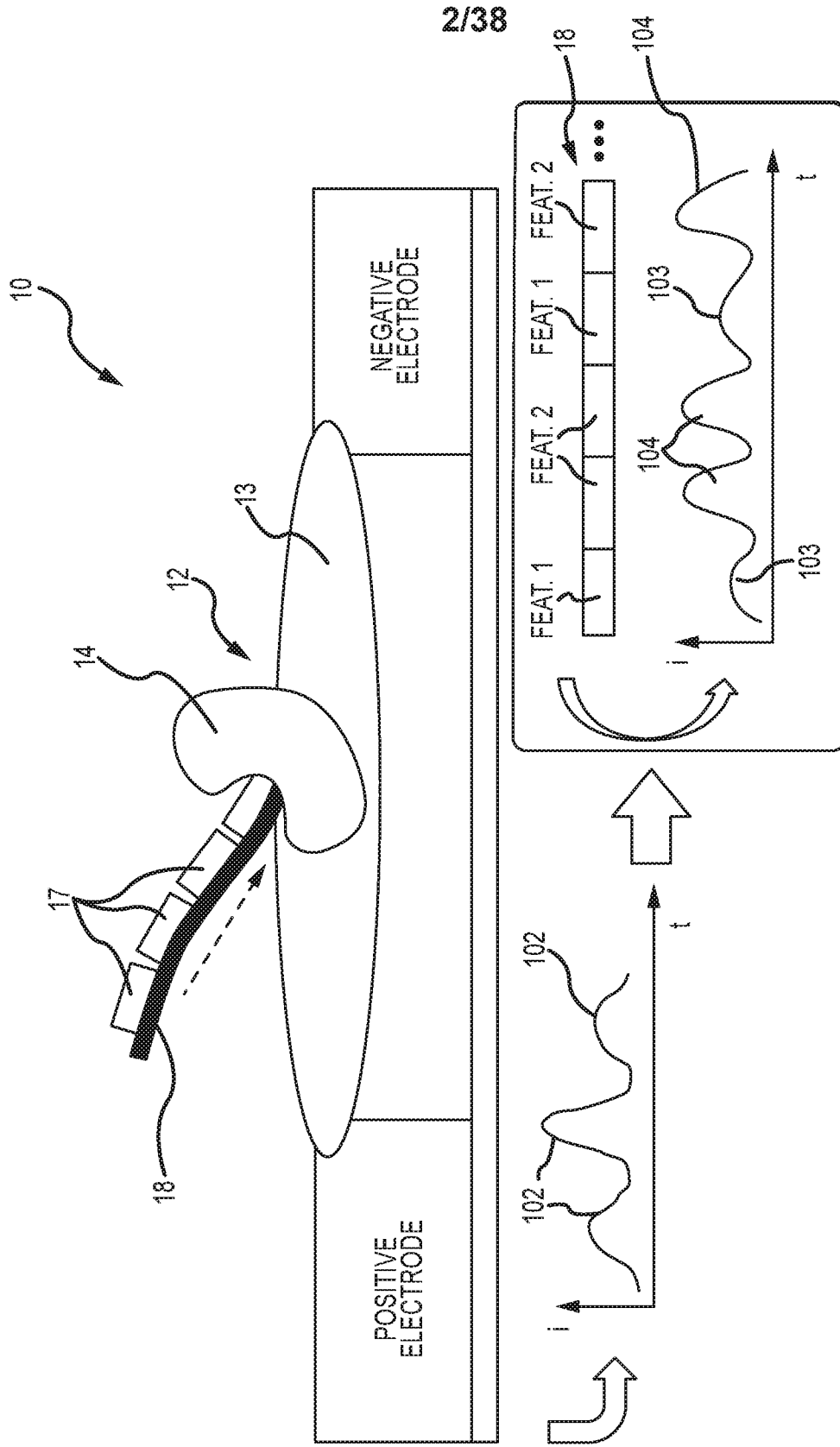


FIG.1-B

3/38

PROGRESSIVE ENZYME EXAMPLES

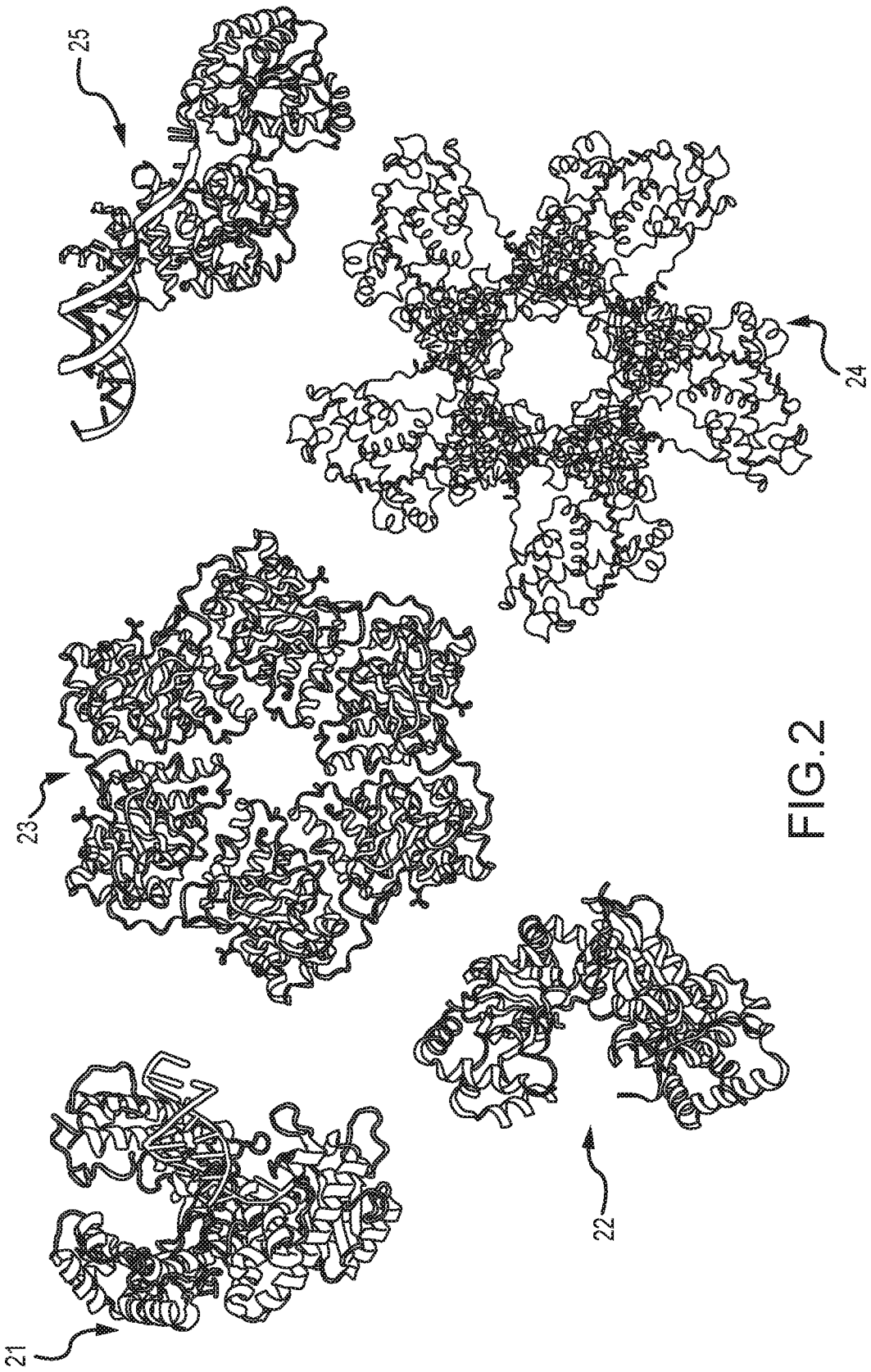


FIG.2

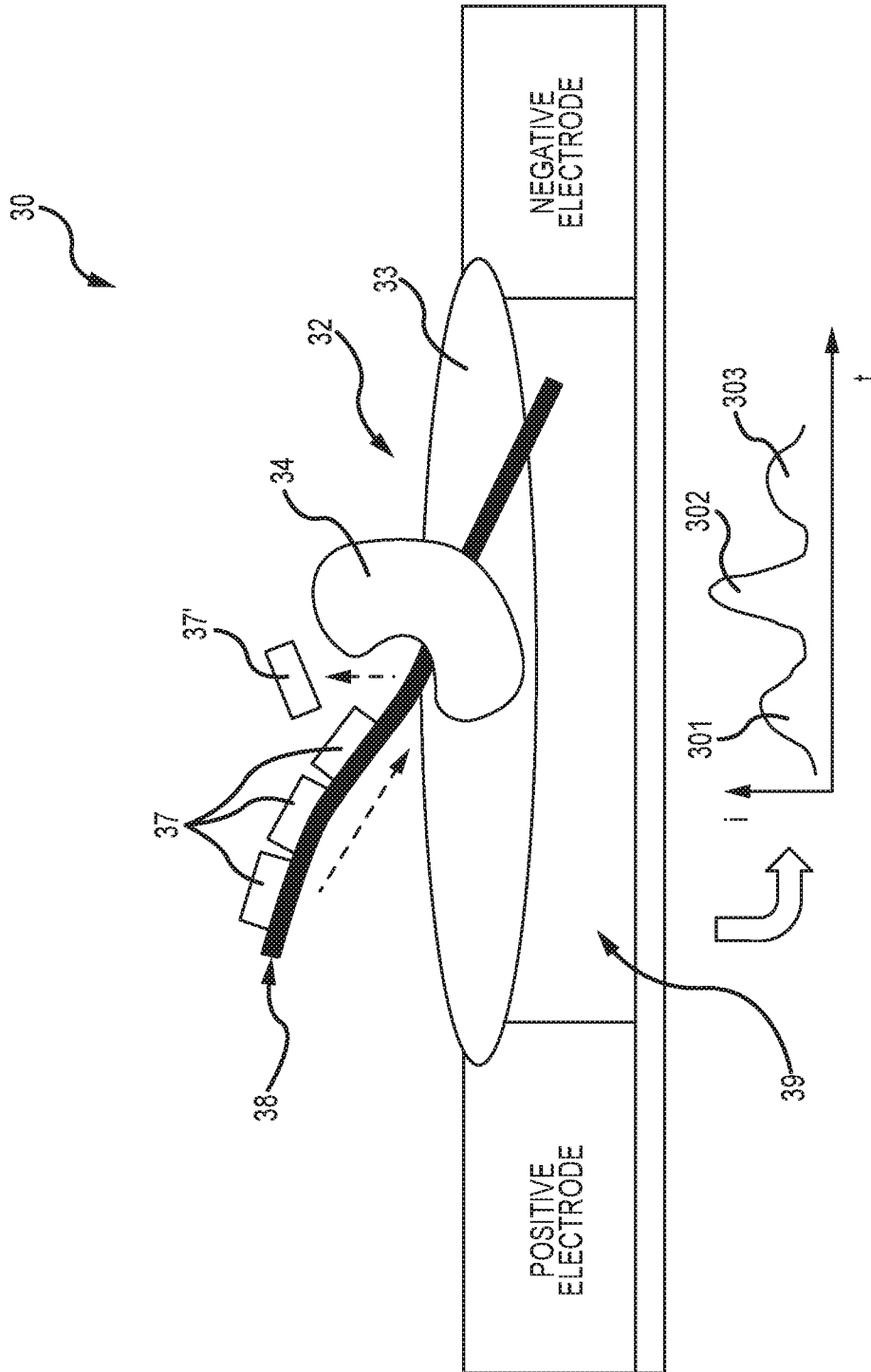


FIG.3-A

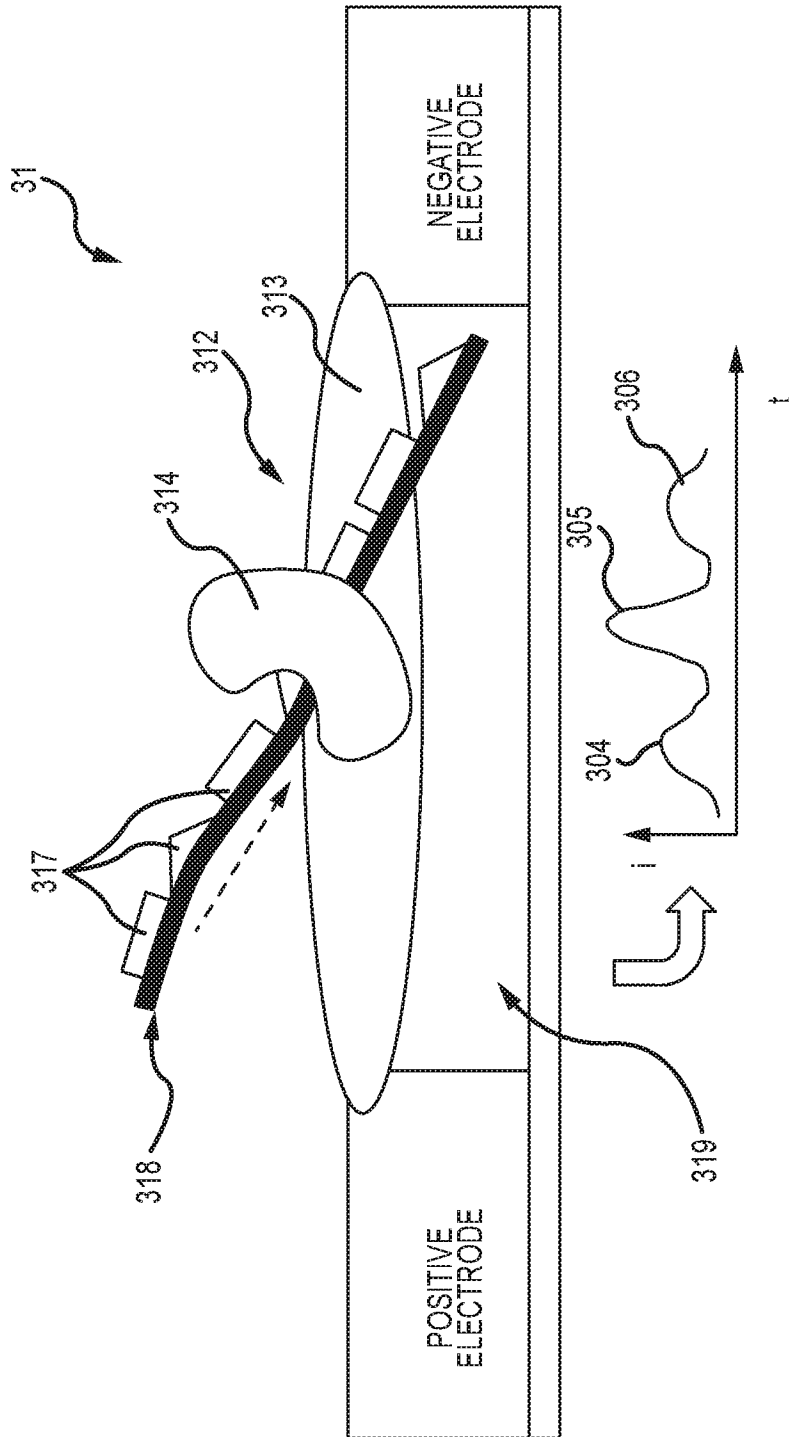


FIG.3-B

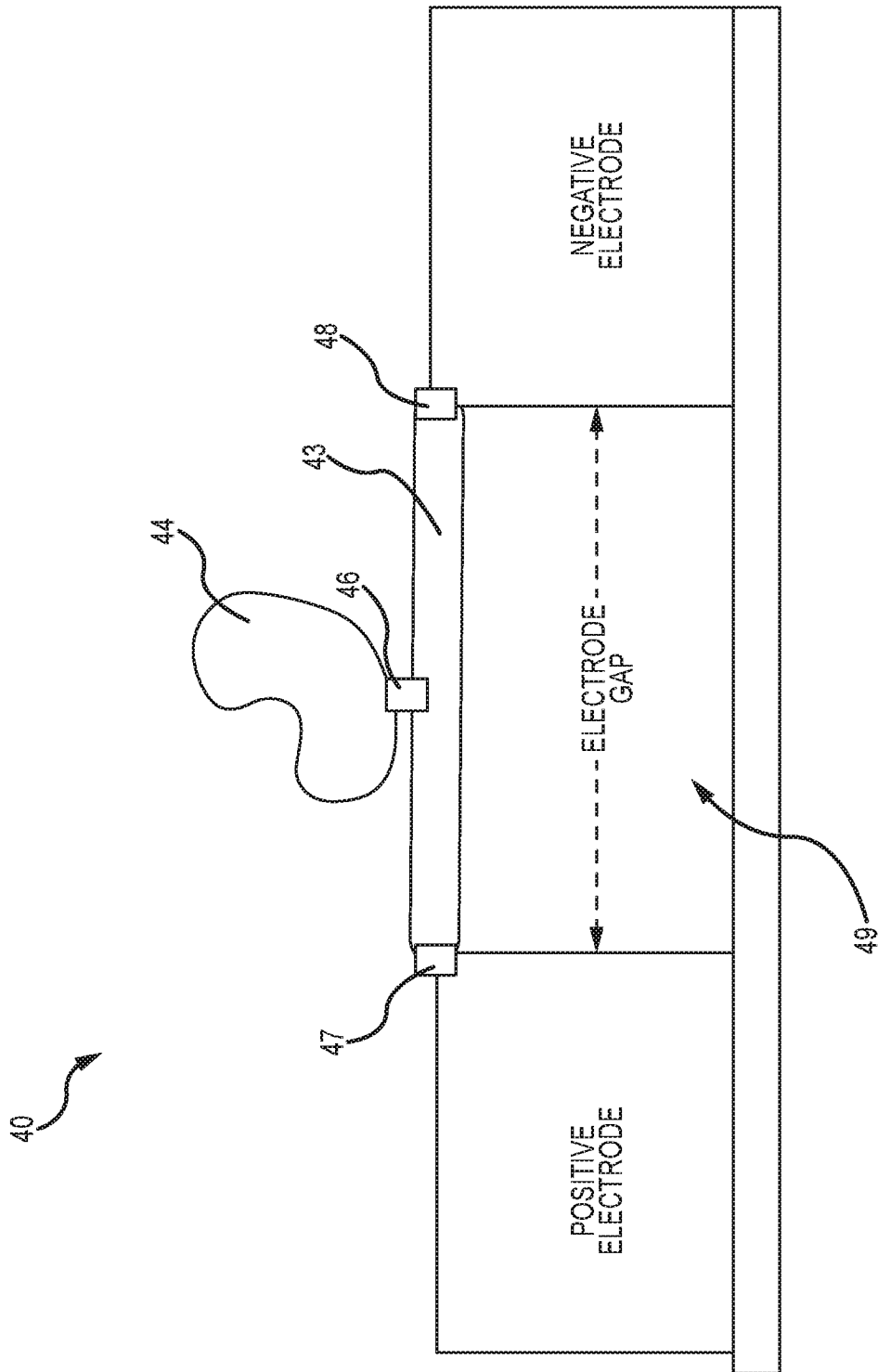


FIG.4-A

7/38

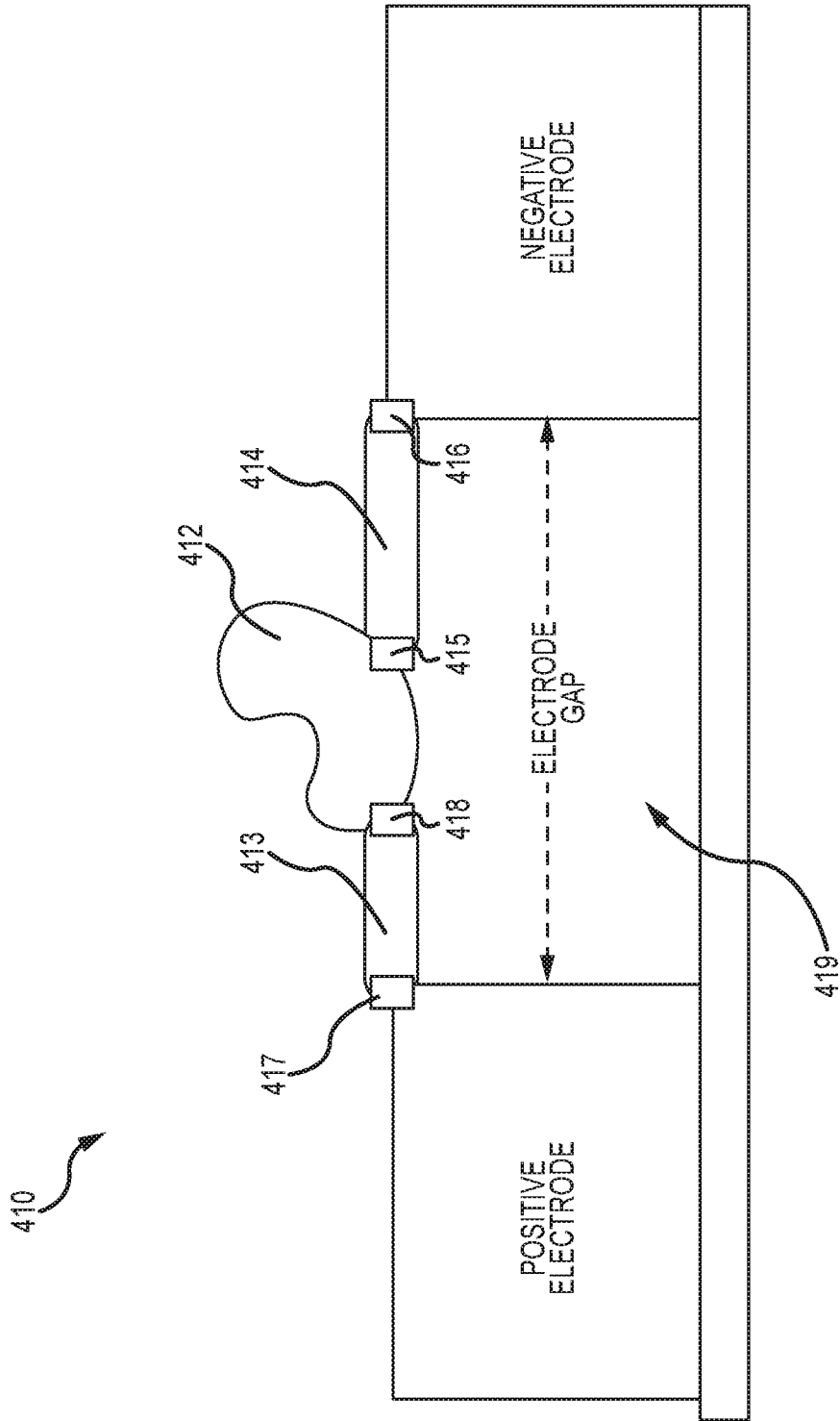


FIG.4-B

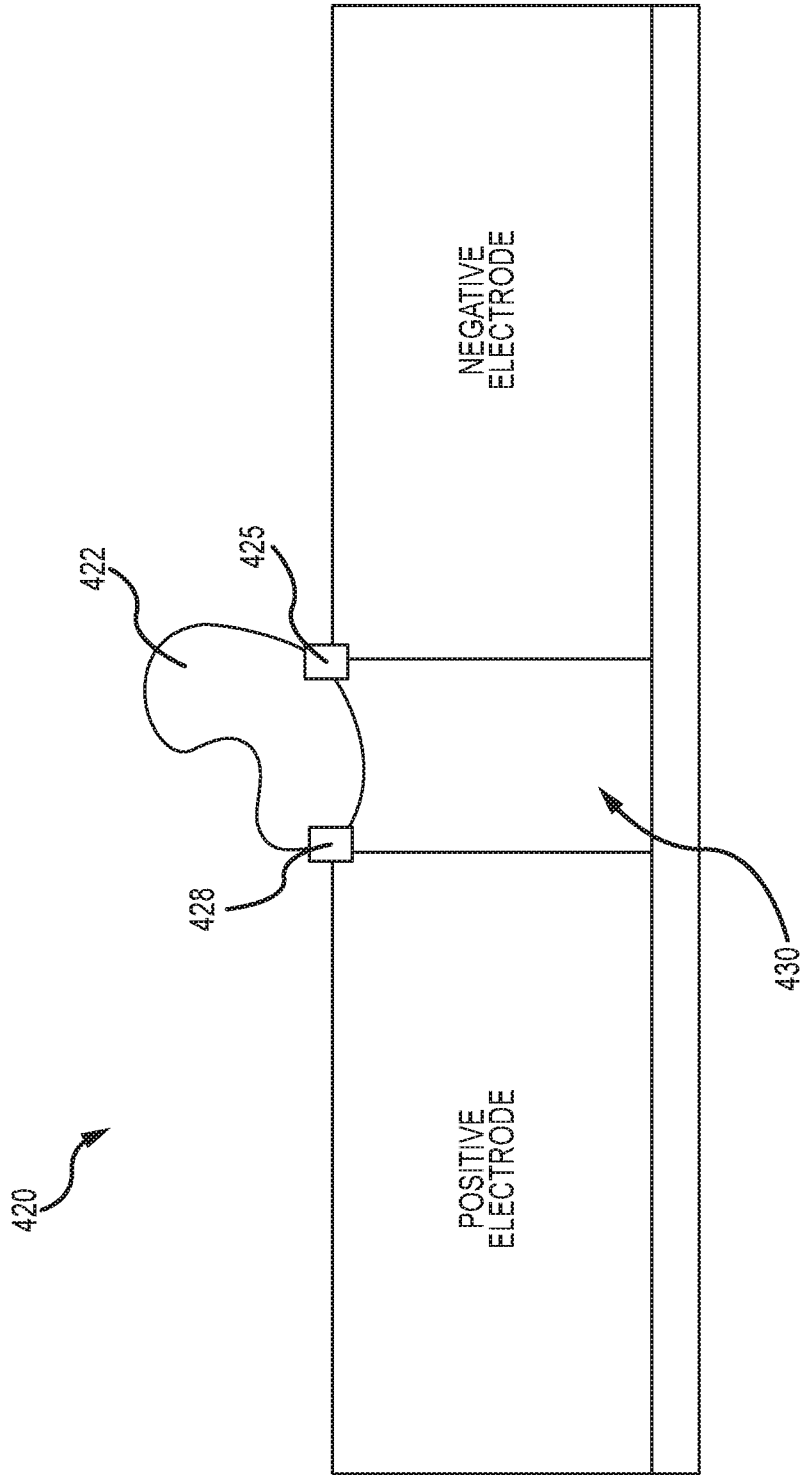


FIG.4-C

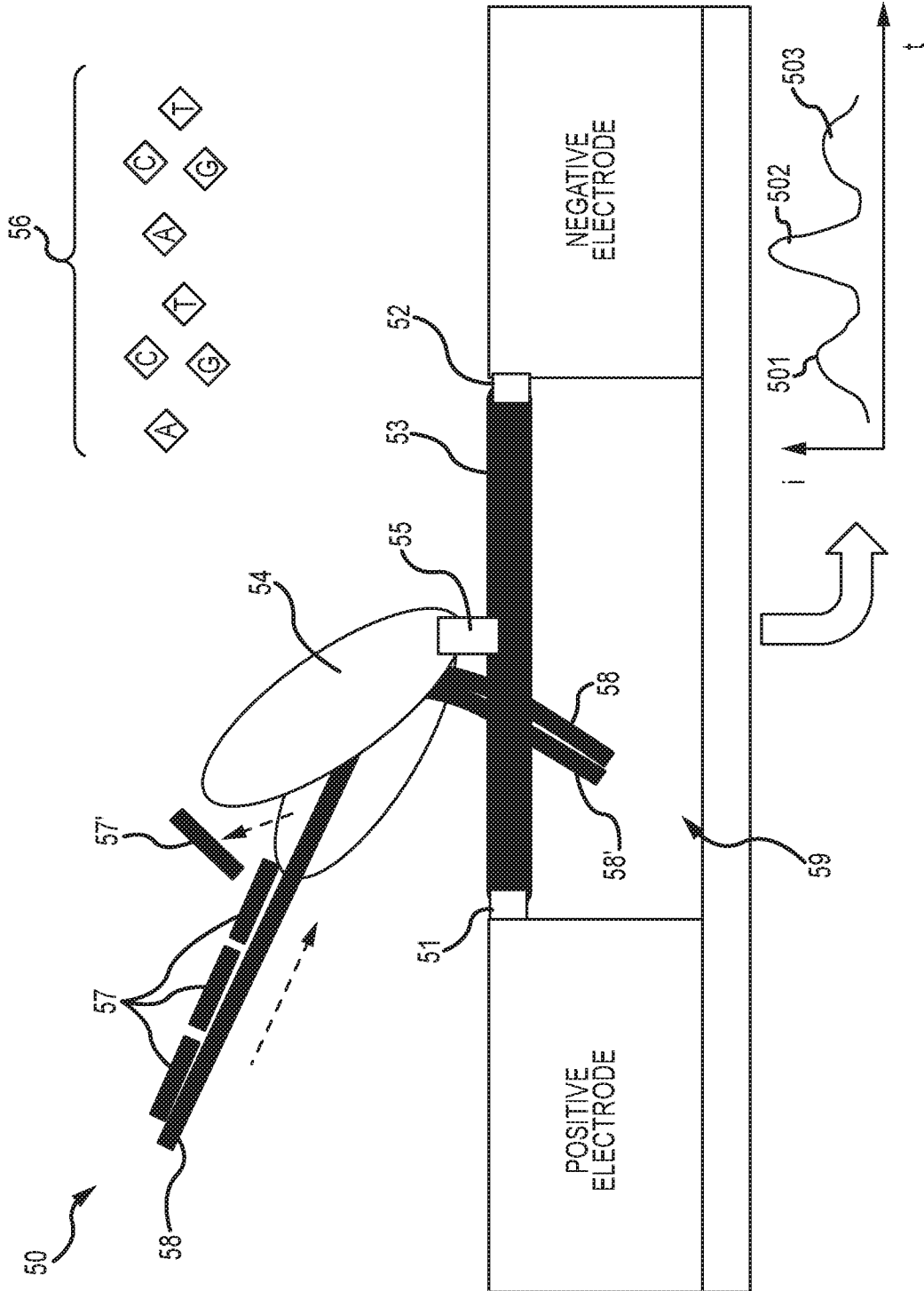


FIG.5

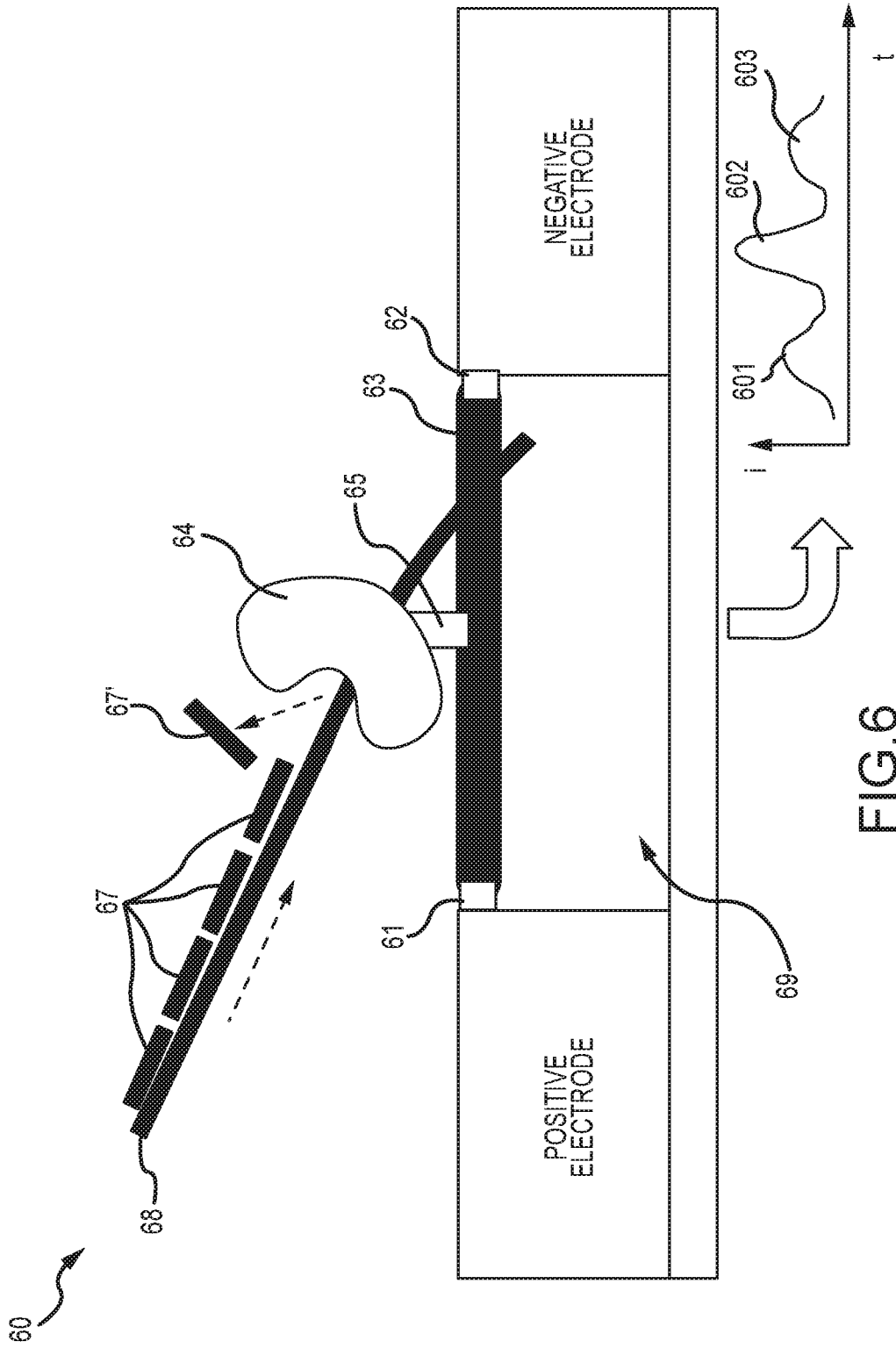


FIG.6

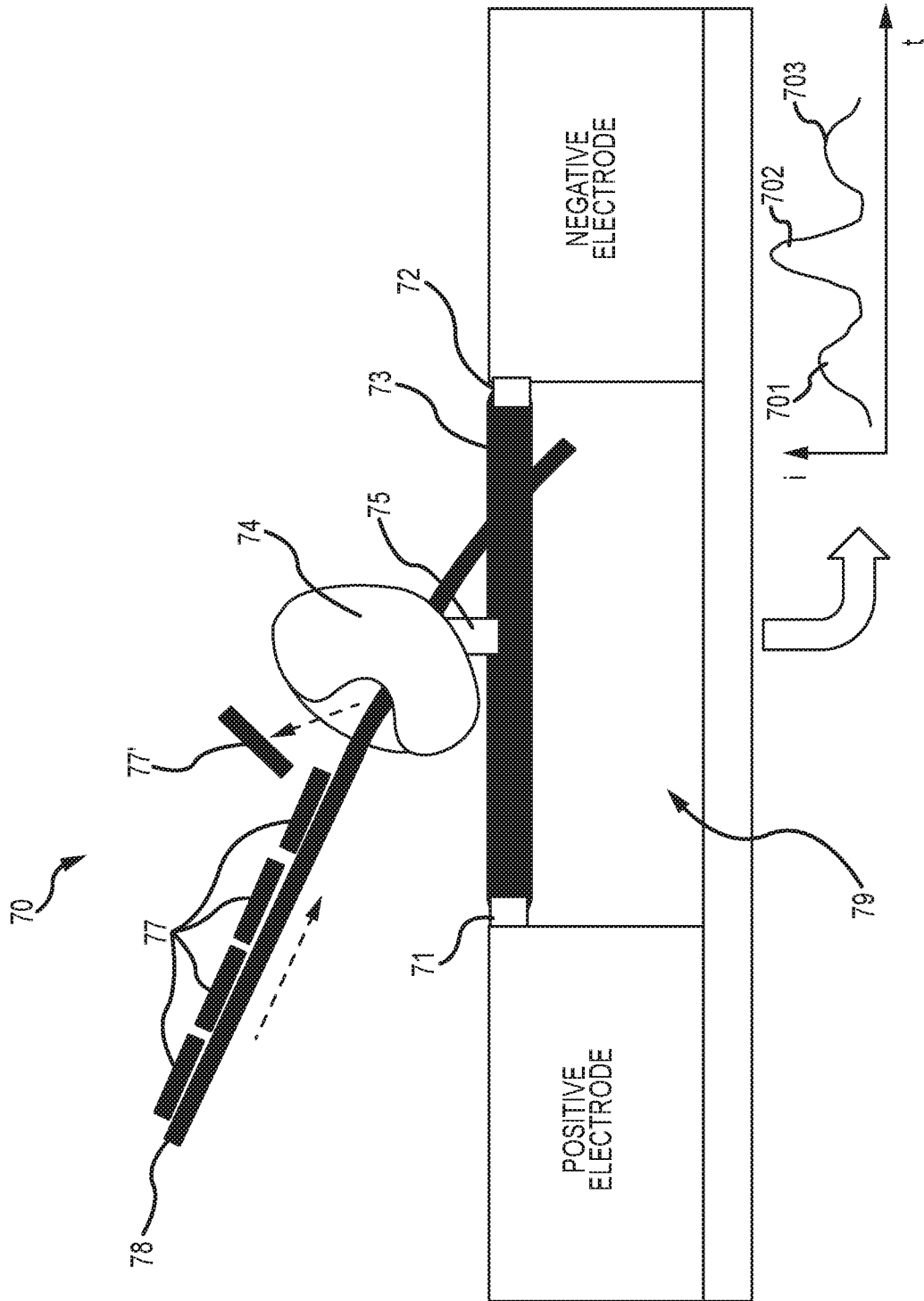


FIG.7

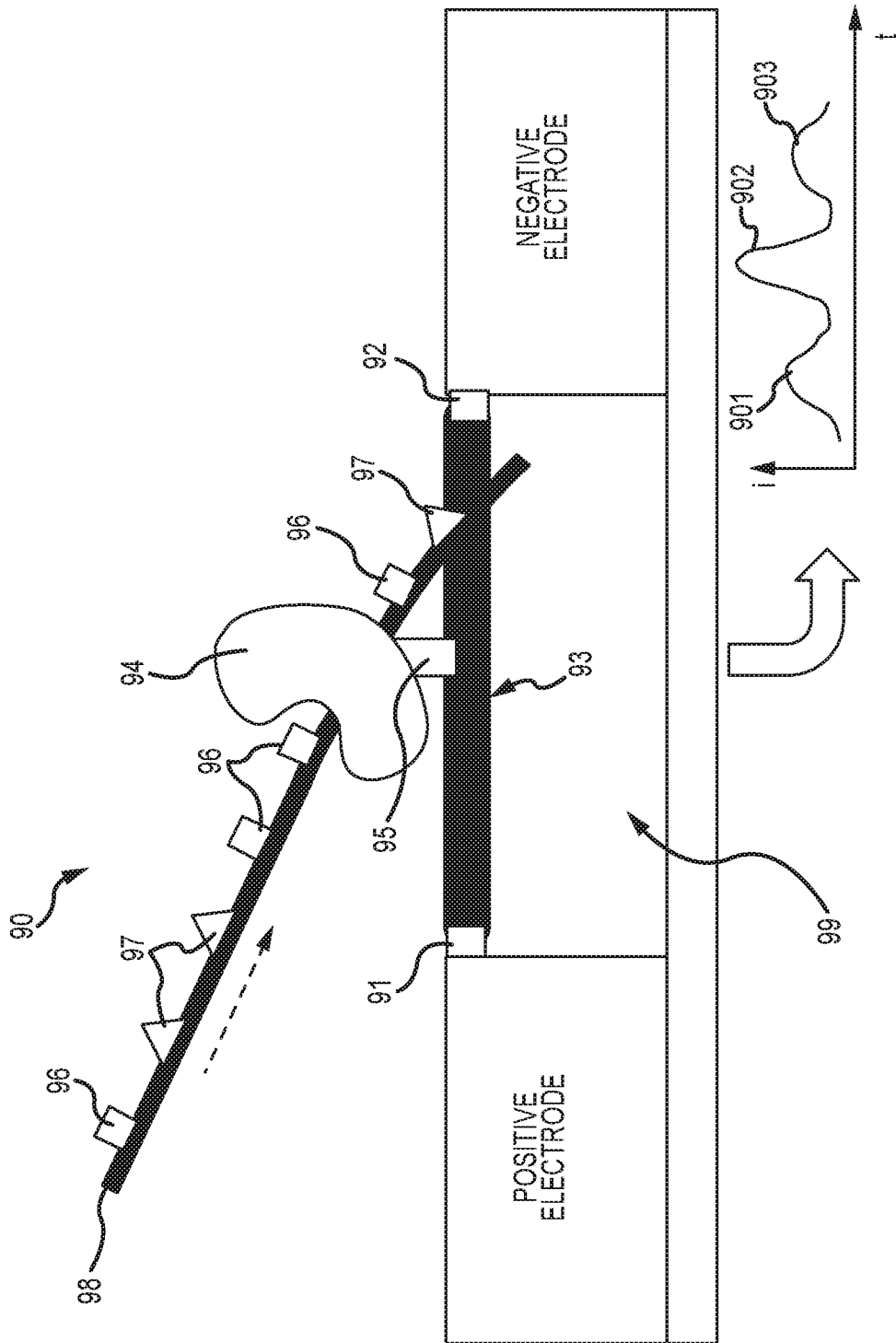


FIG.9

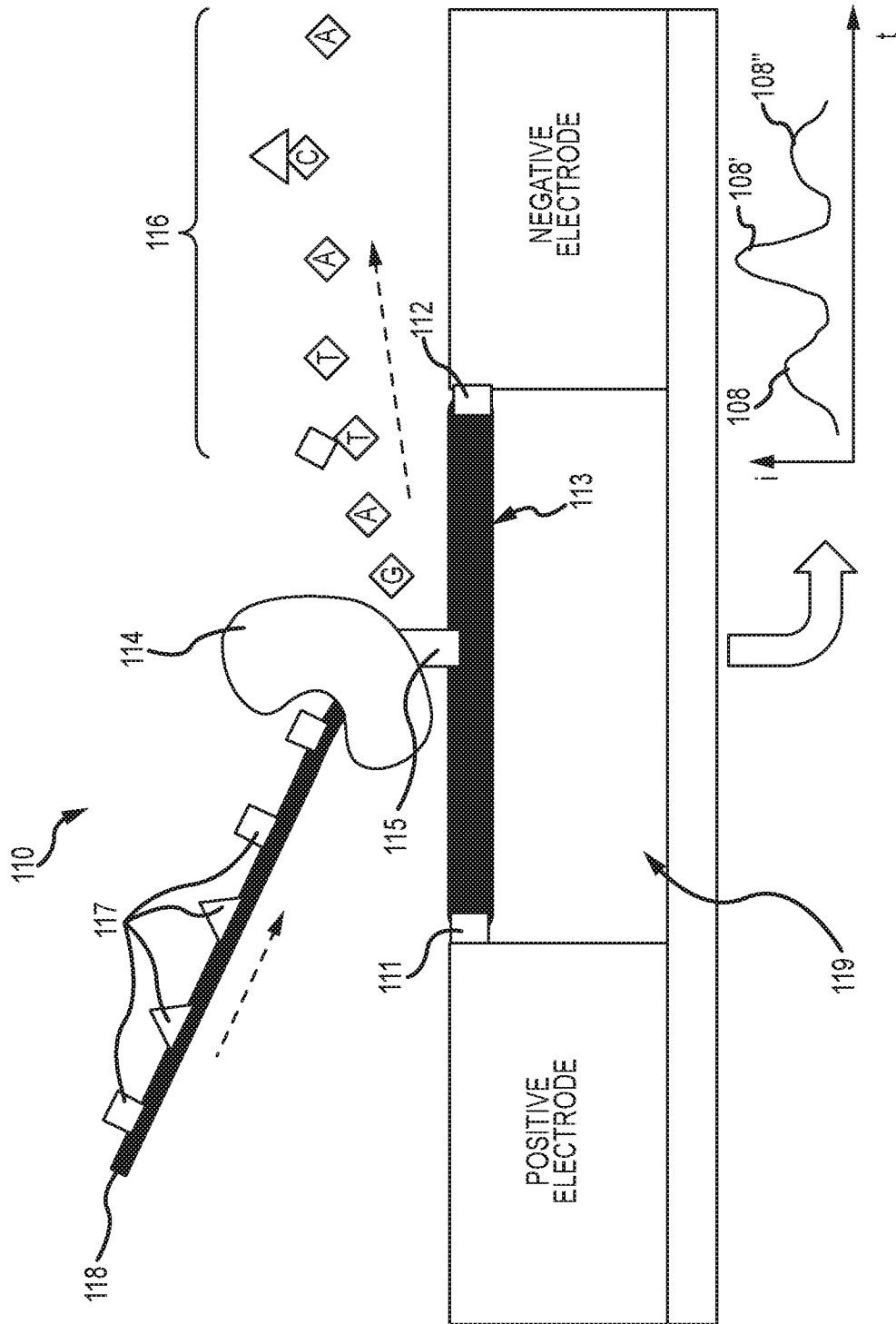


FIG.11

16/38

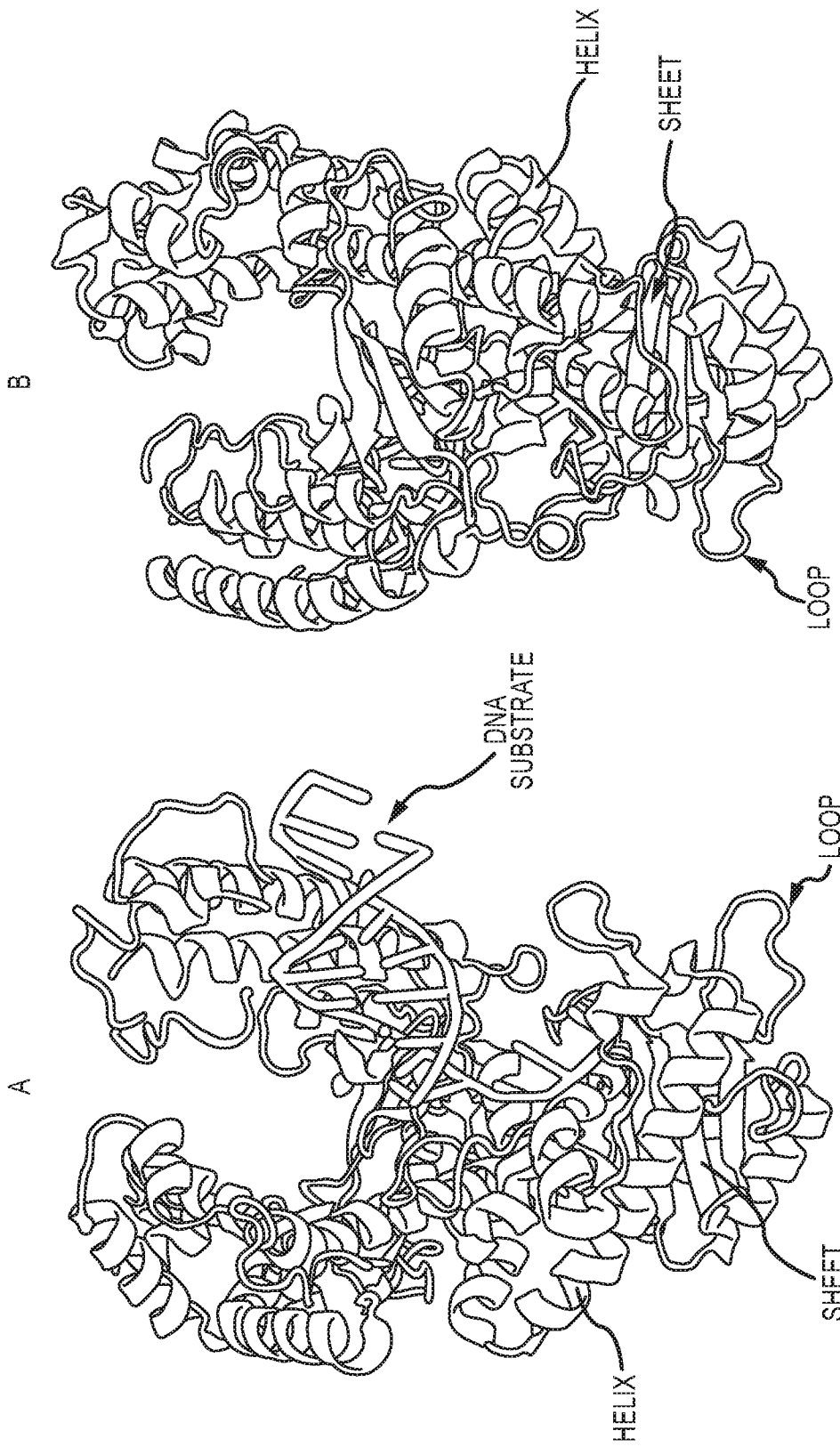


FIG.12-A

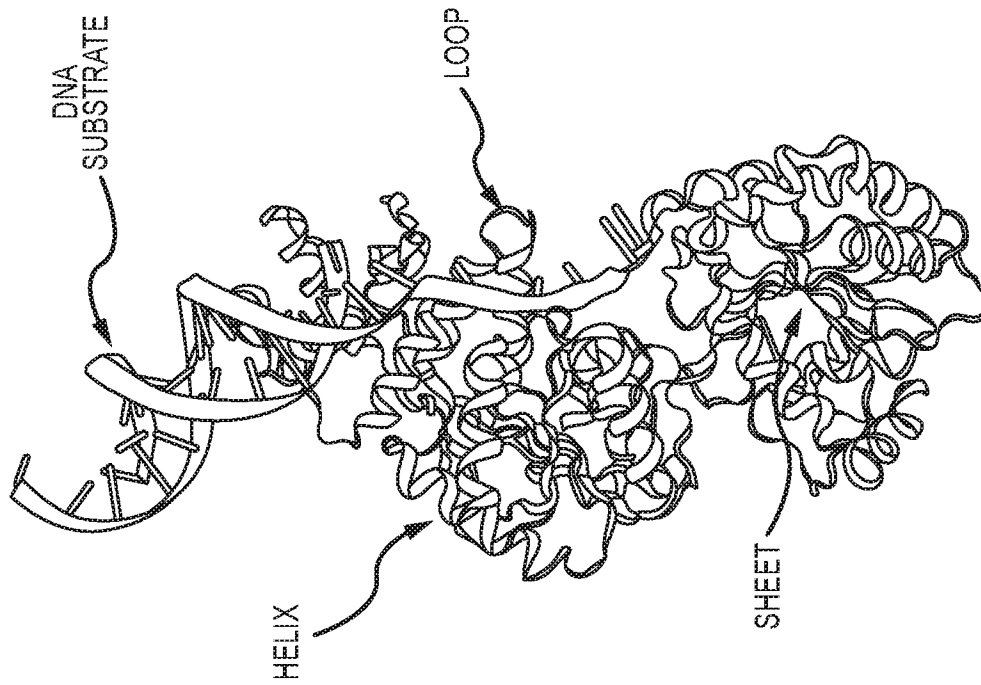


FIG.12-B

18/38

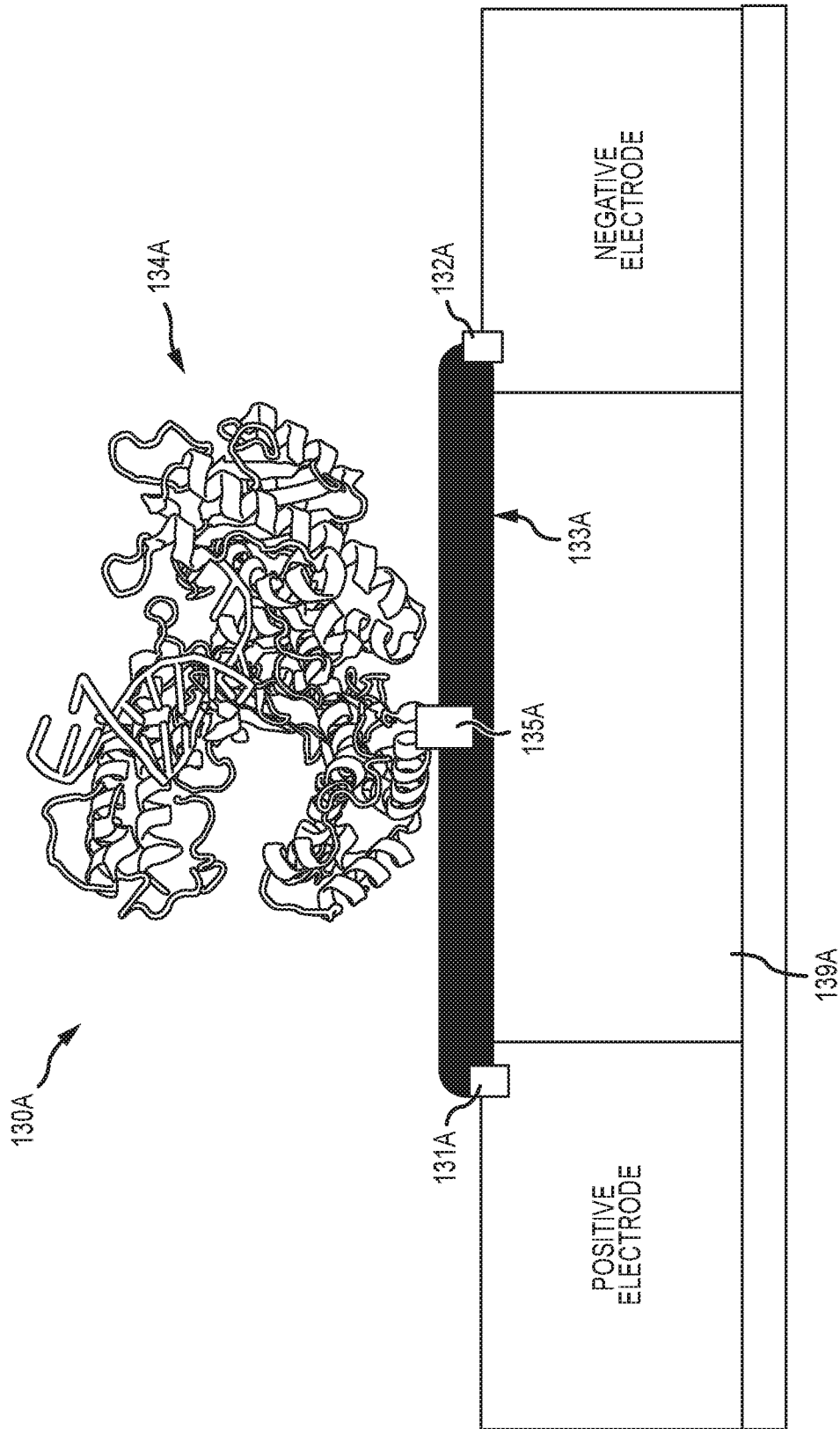


FIG.13-A

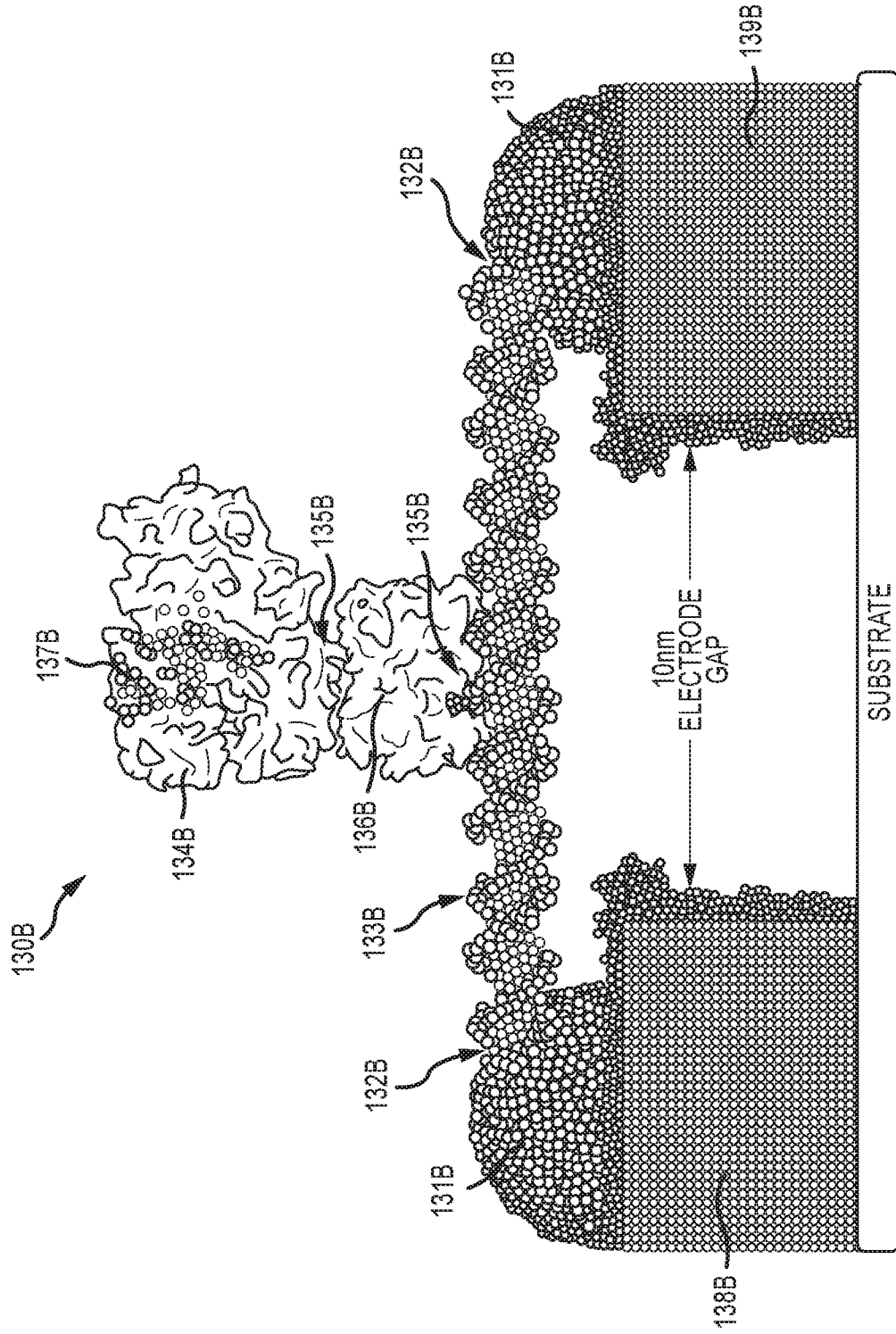


FIG.13-B

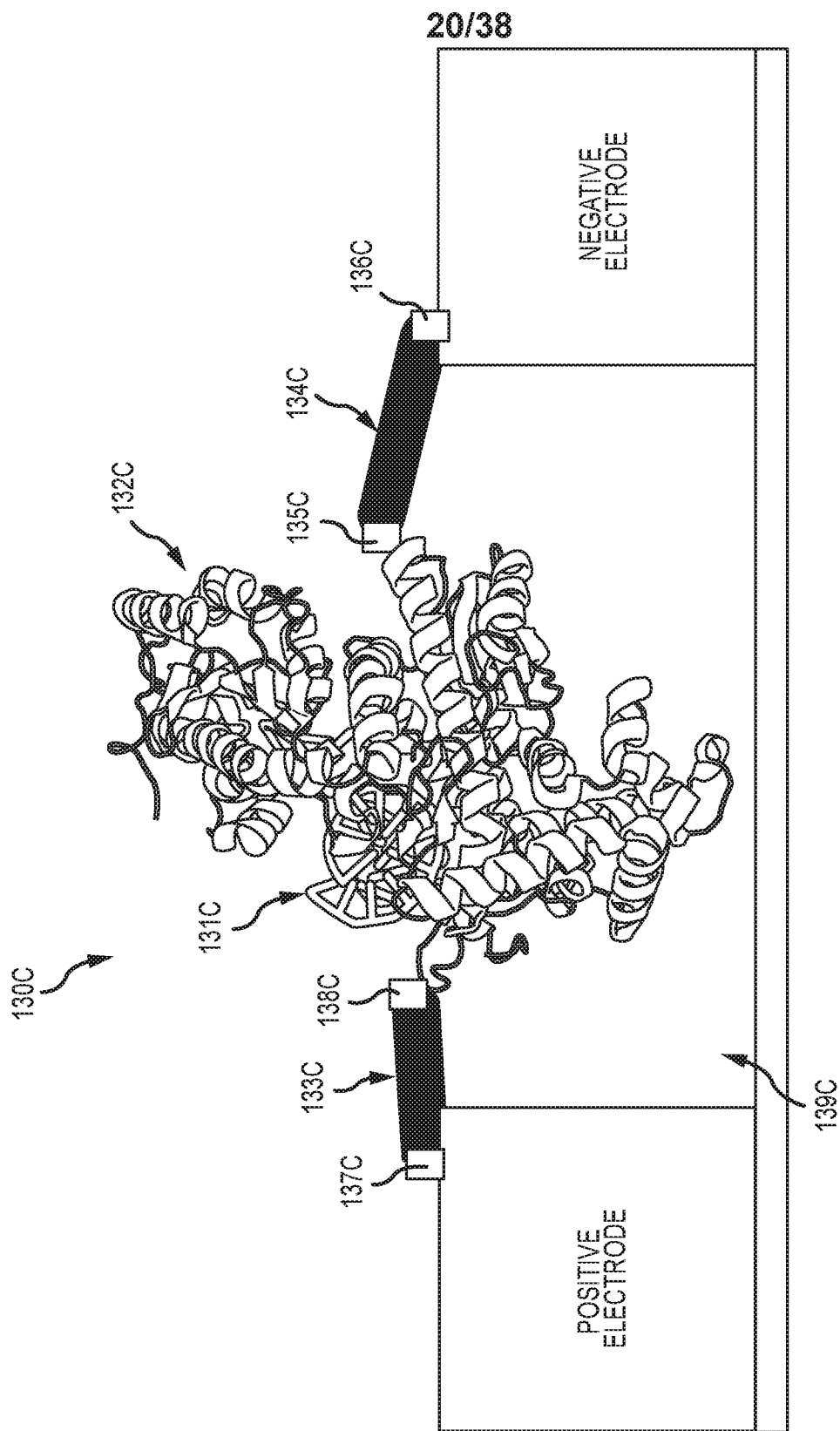


FIG.13-C

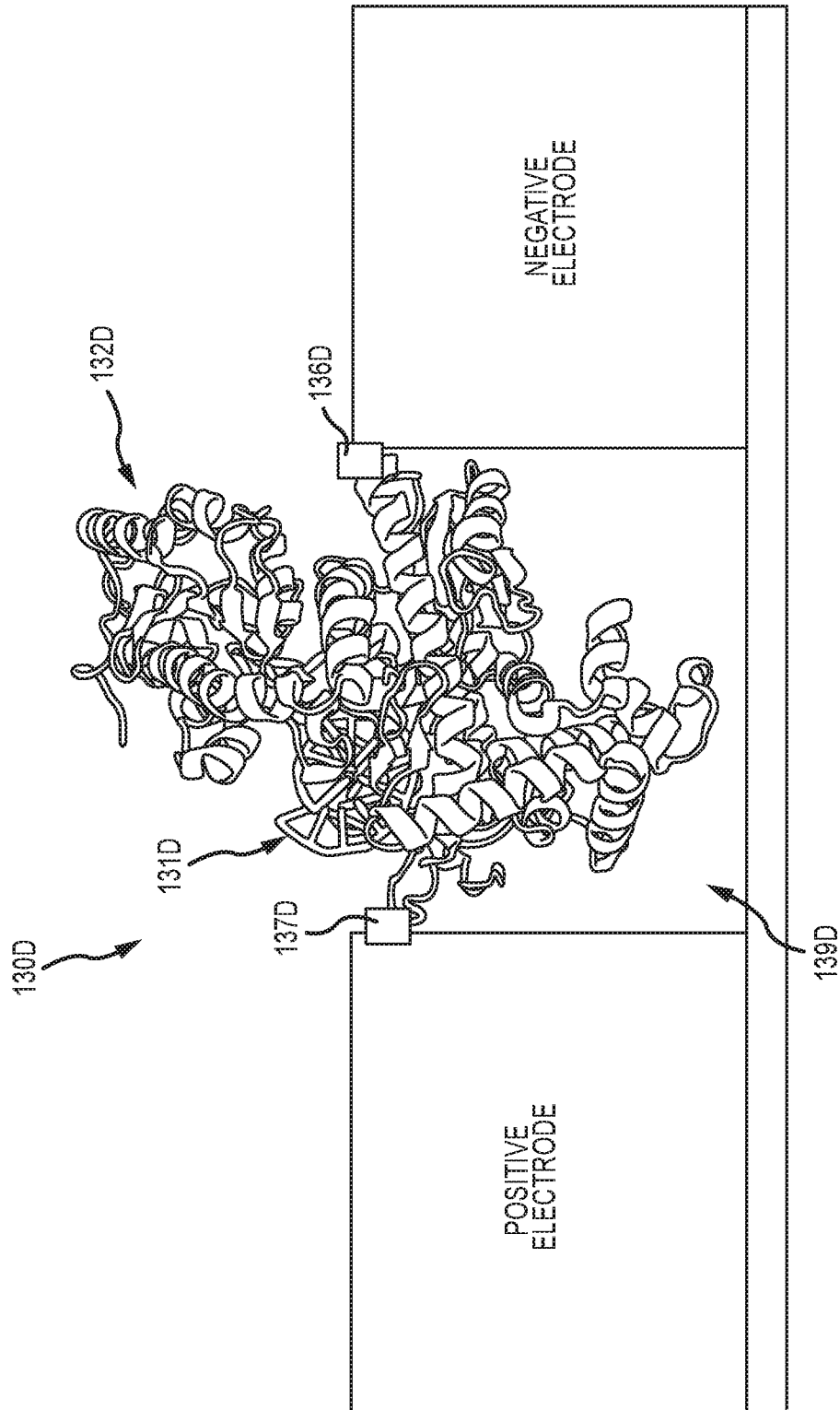


FIG. 13-D

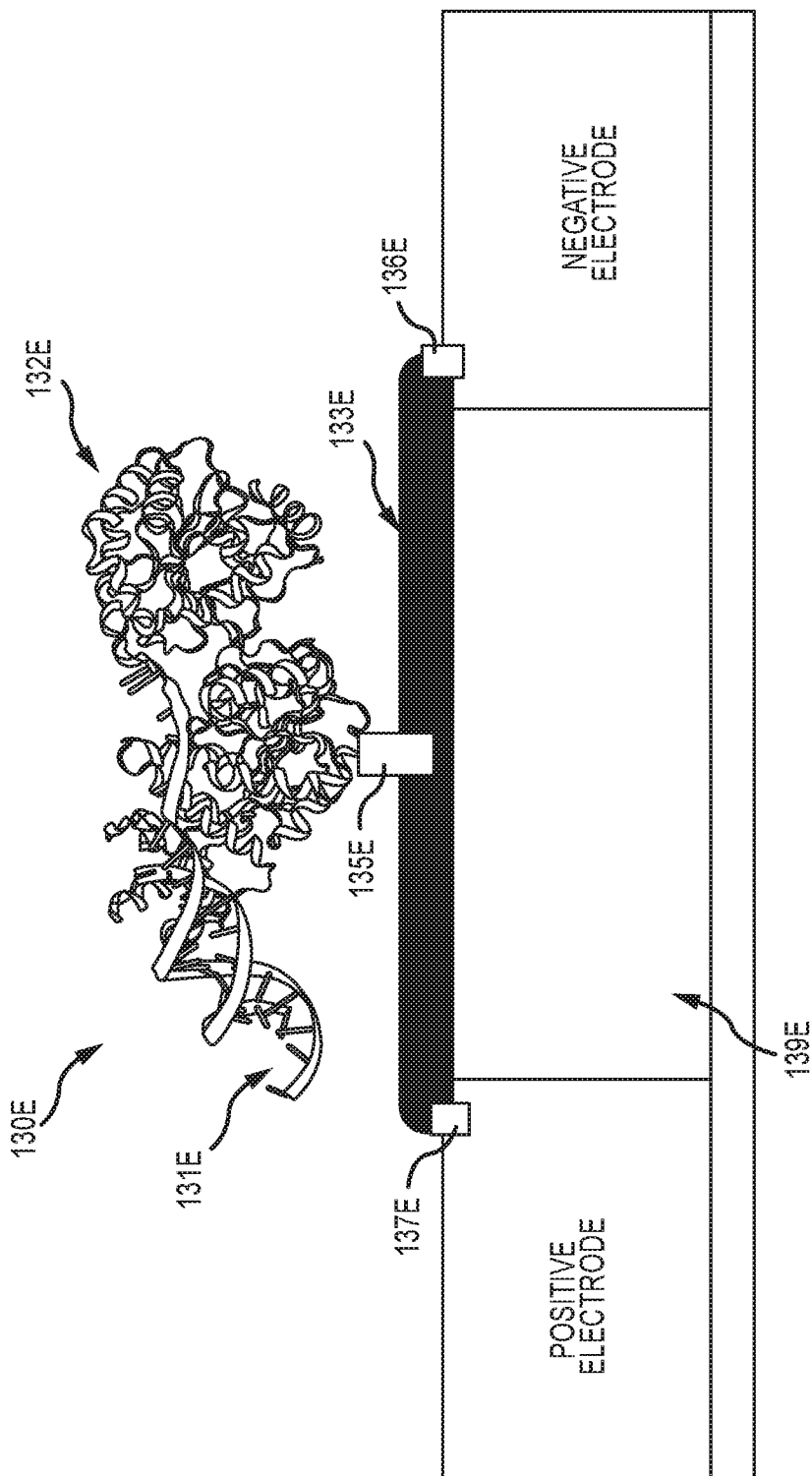


FIG. 13-E

DNA PHYSICAL STRUCTURES FOR DIGITAL DATA STORAGE

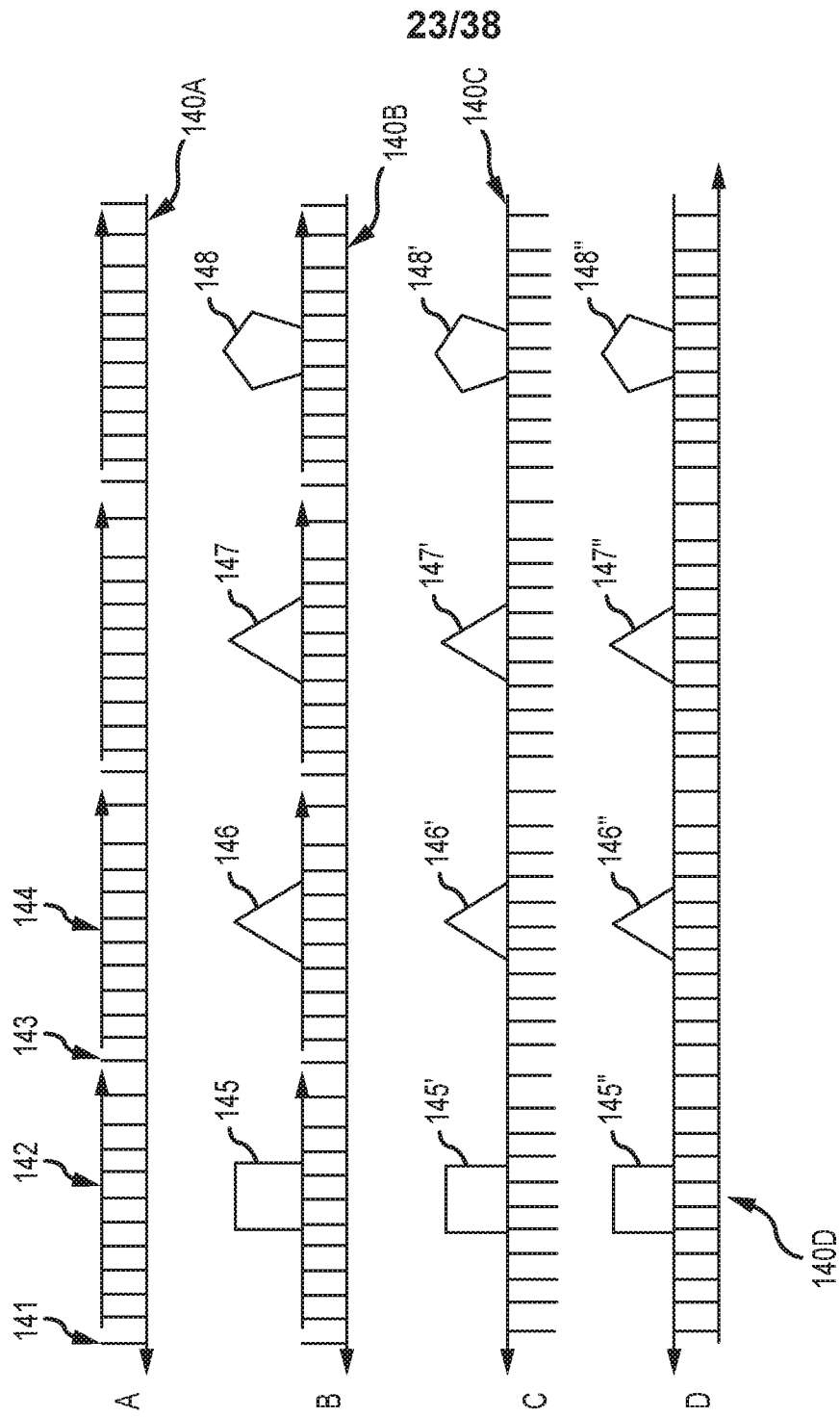


FIG. 14-A

24/38

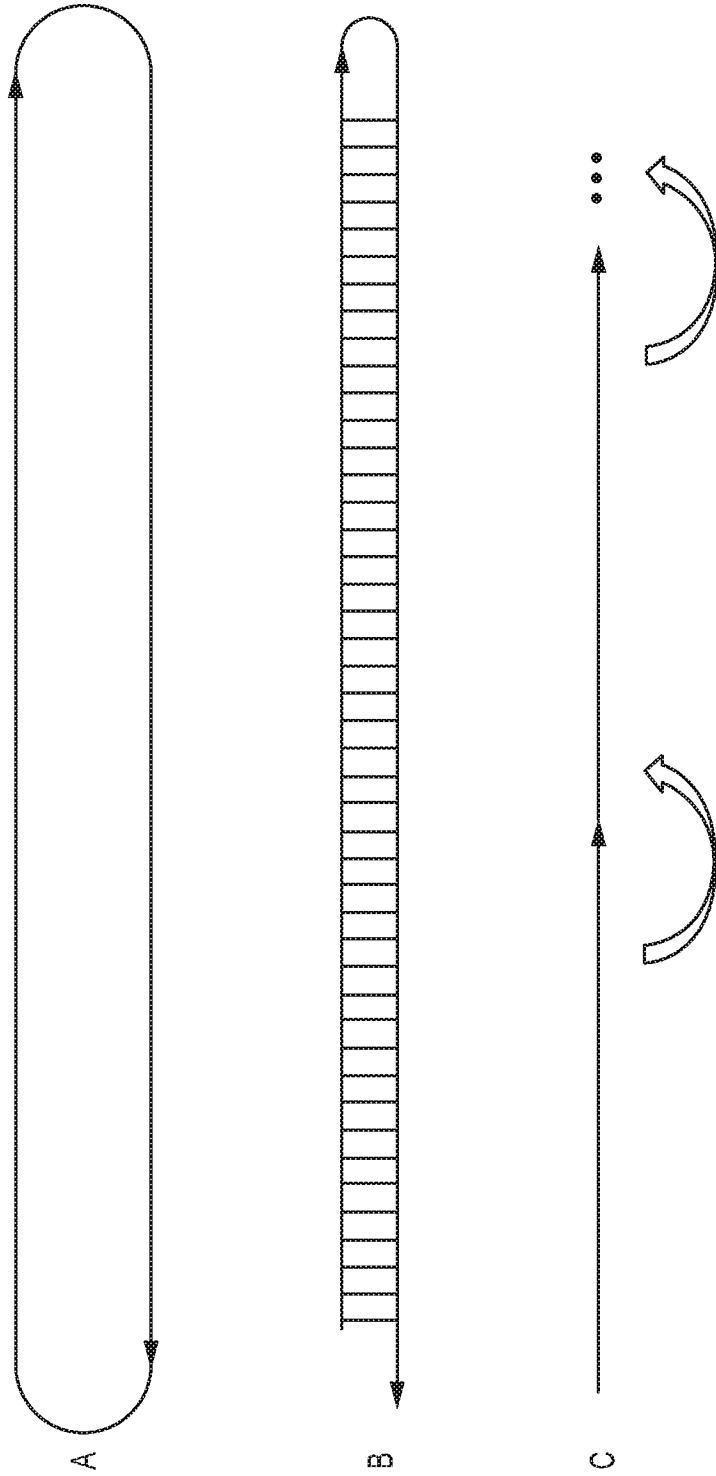


FIG.14-B

DATA STORAGE DNA LOGICAL STRUCTURES

25/38

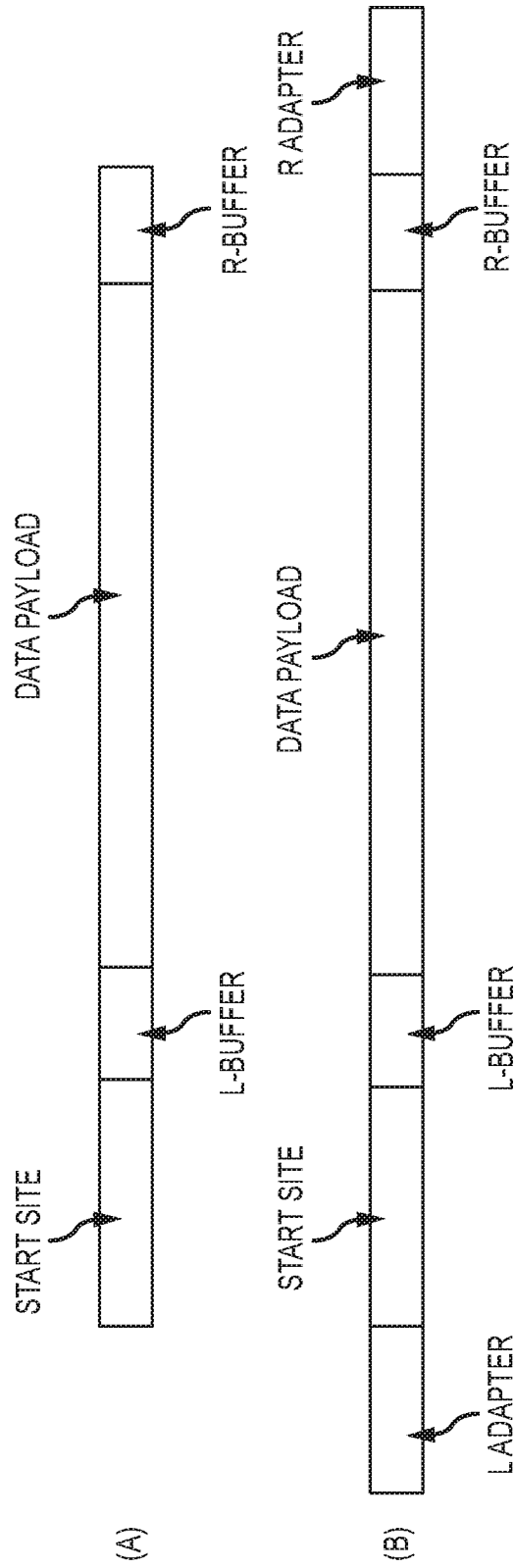


FIG.15

EXAMPLE BINARY DATA PAYLOAD 1010100110011100

EXAMPLE BINARY ENCODING SCHEMES (BES)

BES1	BES2	BES3
0 → F1	00 → F1	0 → F1
1 → F2	01 → F2	1 → F2
	10 → F3	00 → F3
	11 → F4	

BES1 1 0 1 0 1 0 0 1 1 0 0 1 1 1 0 0
 F2F1F2F1F2F1F1F2F2F1F1F2F2F2F1F1

BES2 1010100110011100
 F3F3F3F2F3F2F4F1

BES3 1 0 1 0 1 0 0 1 1 0 0 1 1 1 0 0
 F2F1F2F1F2F3 F2F2F3 F2F2F2F3

FIG.16

27/38

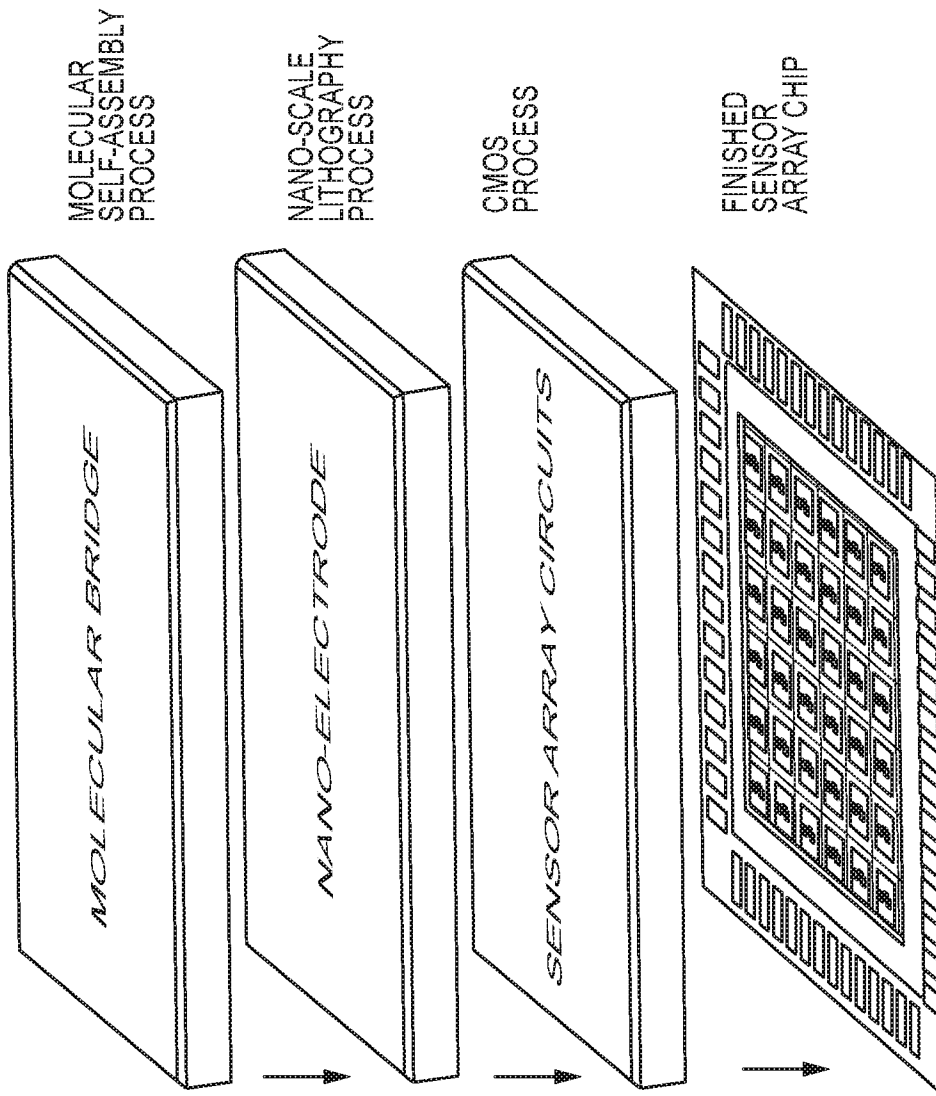


FIG.17-A

CMOS CHIP AND PIXEL ARCHITECTURE

PIXEL ARRAY AND SUPPORTING CHIP ARCHITECTURE

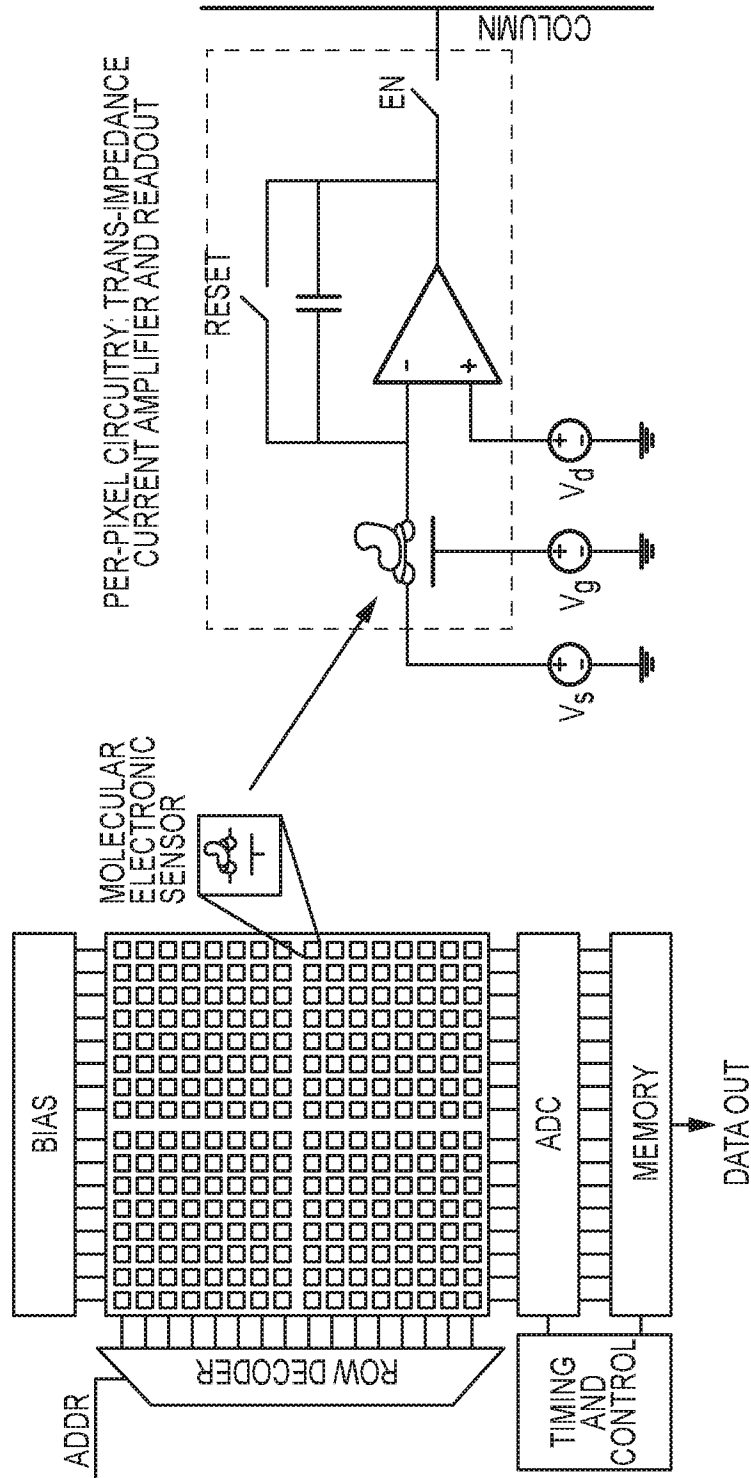


FIG.17-B

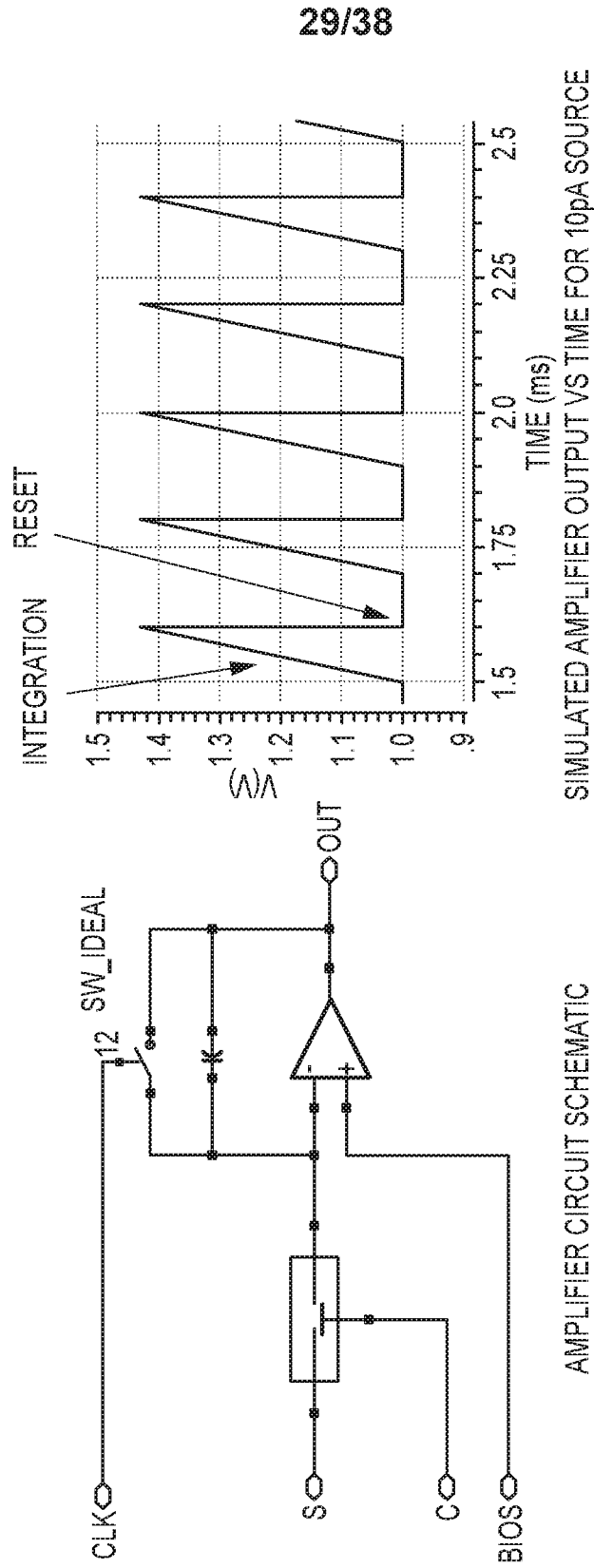


FIG.18

30/38

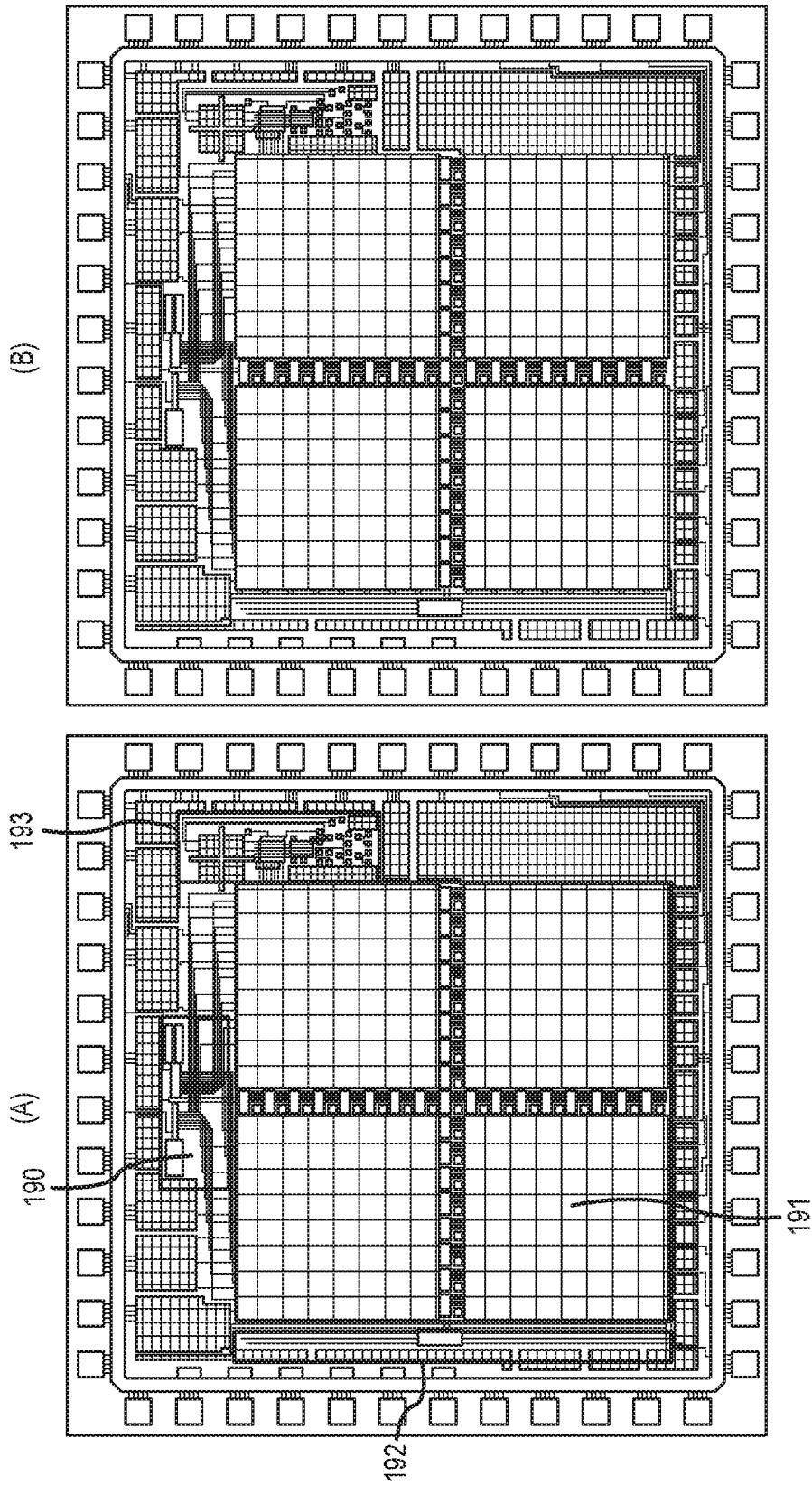


FIG.19

31/38

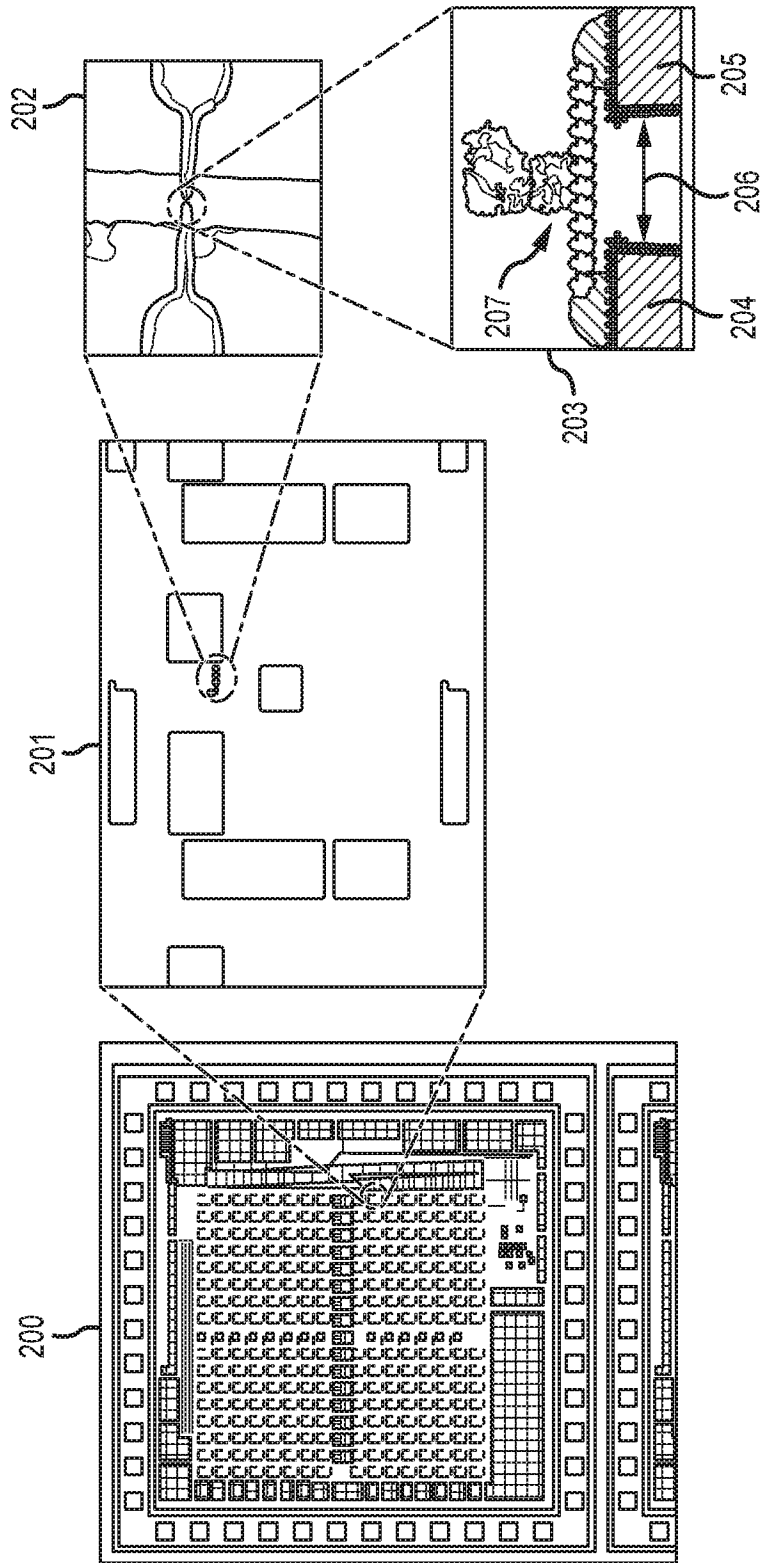


FIG. 20

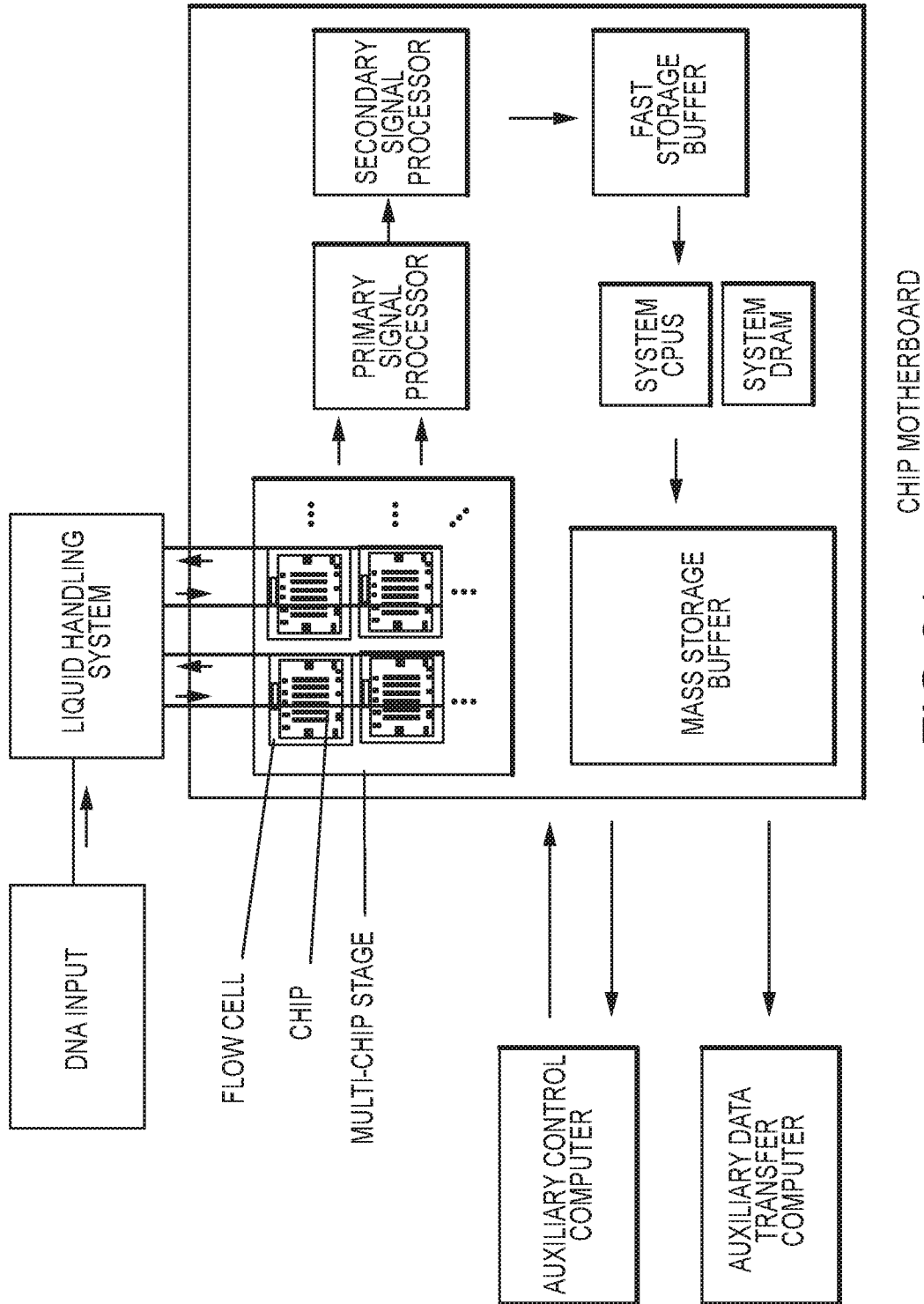


FIG.21

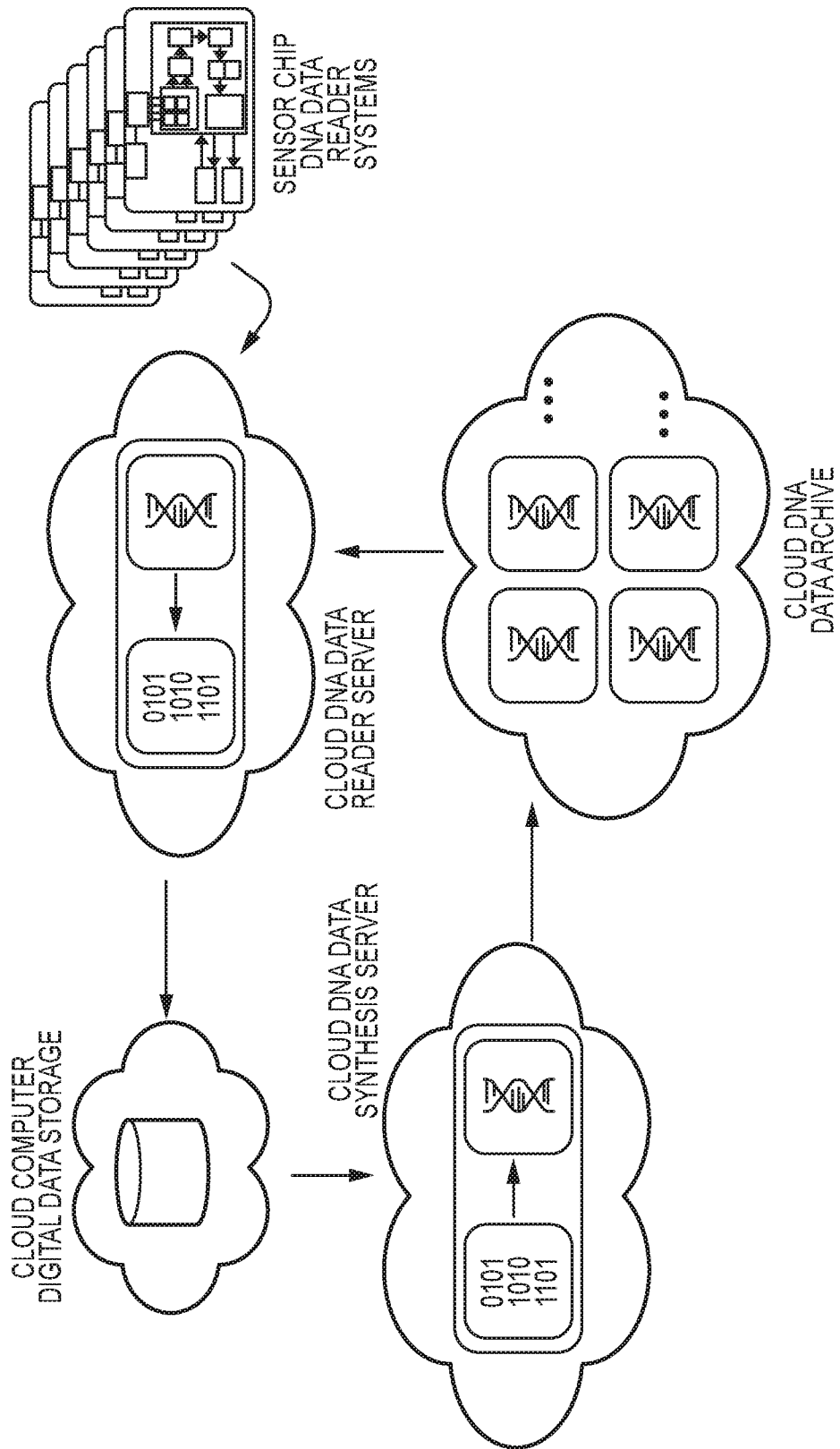


FIG.22

34/38

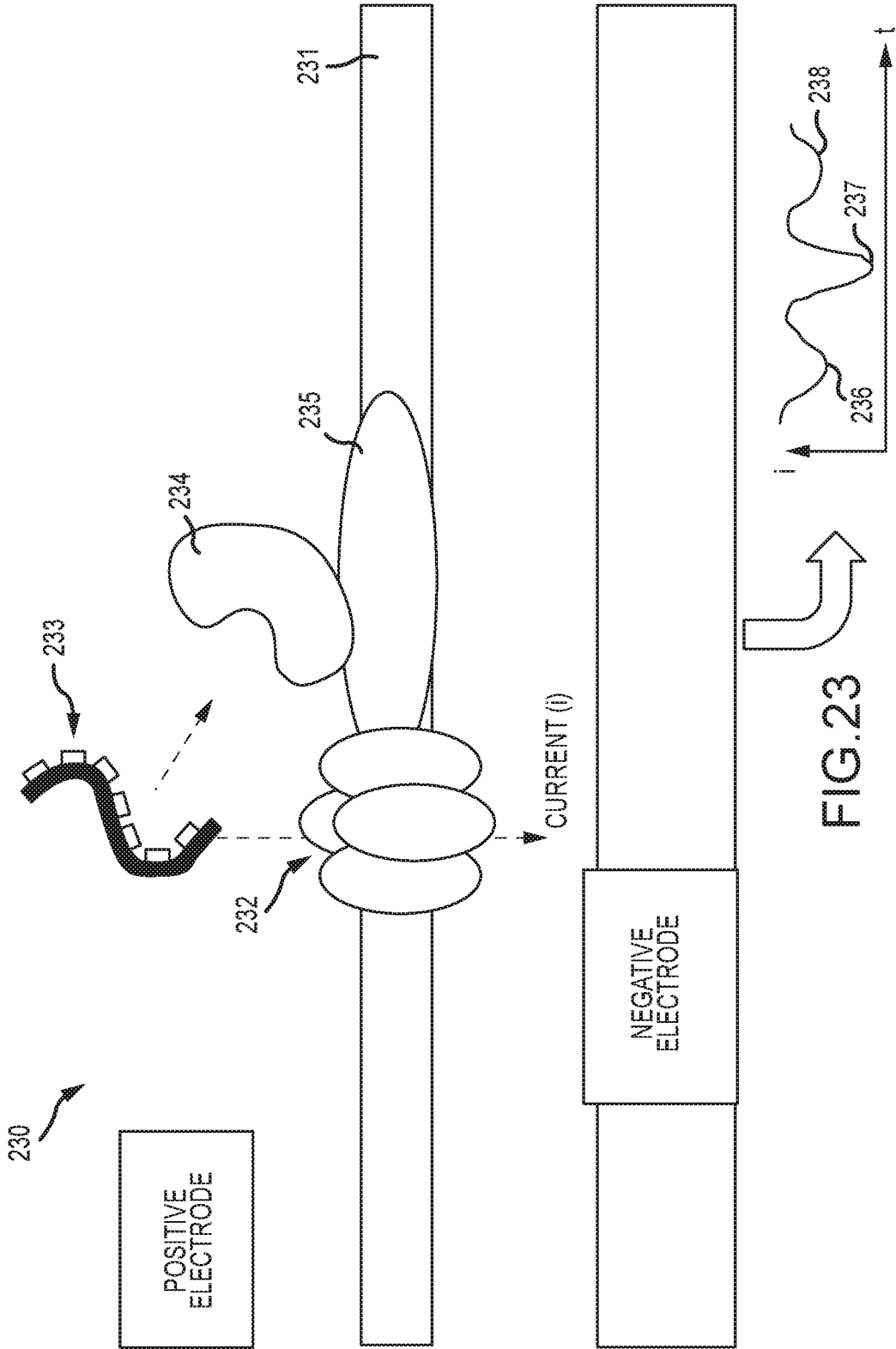


FIG.23

35/38

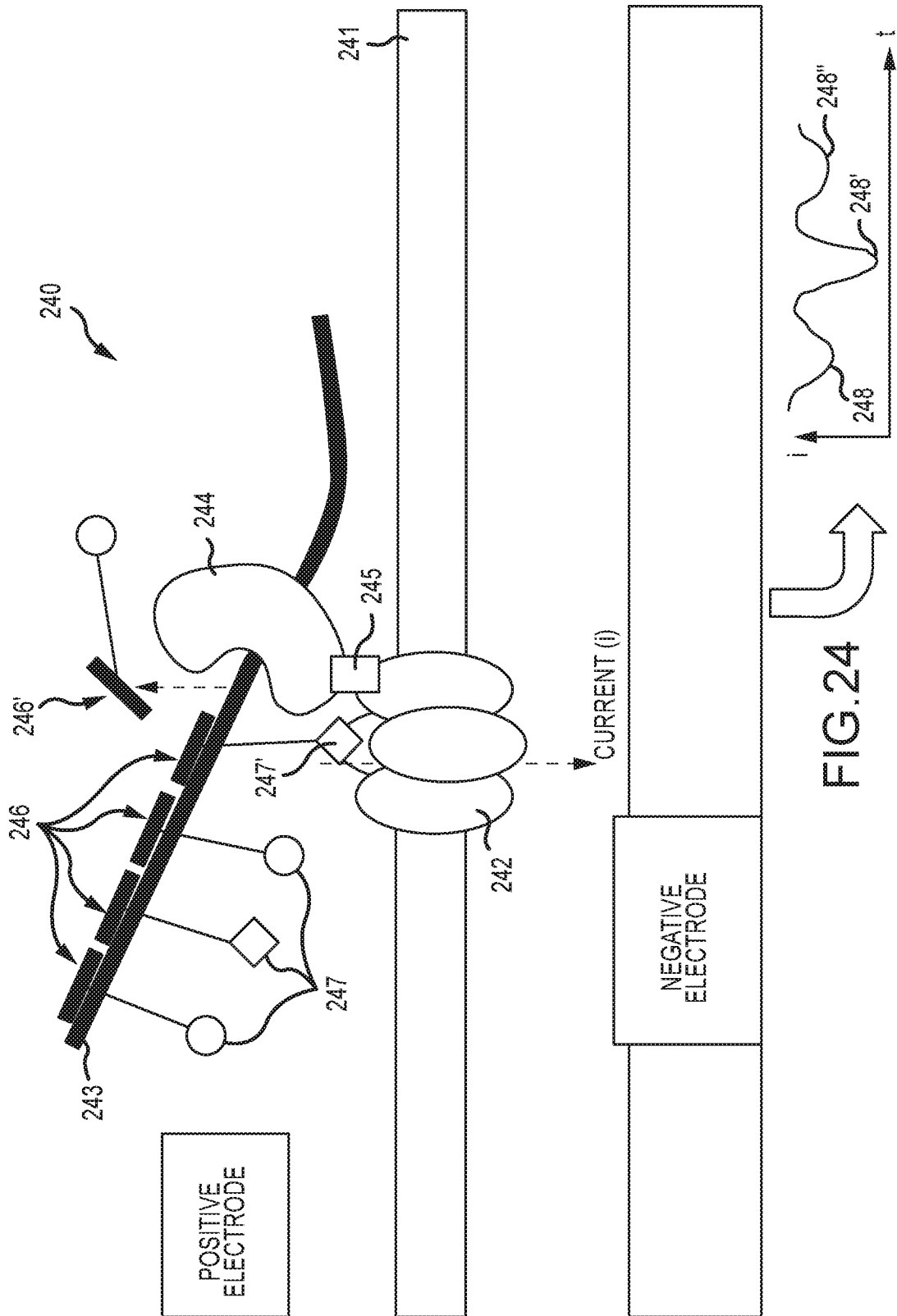


FIG.24

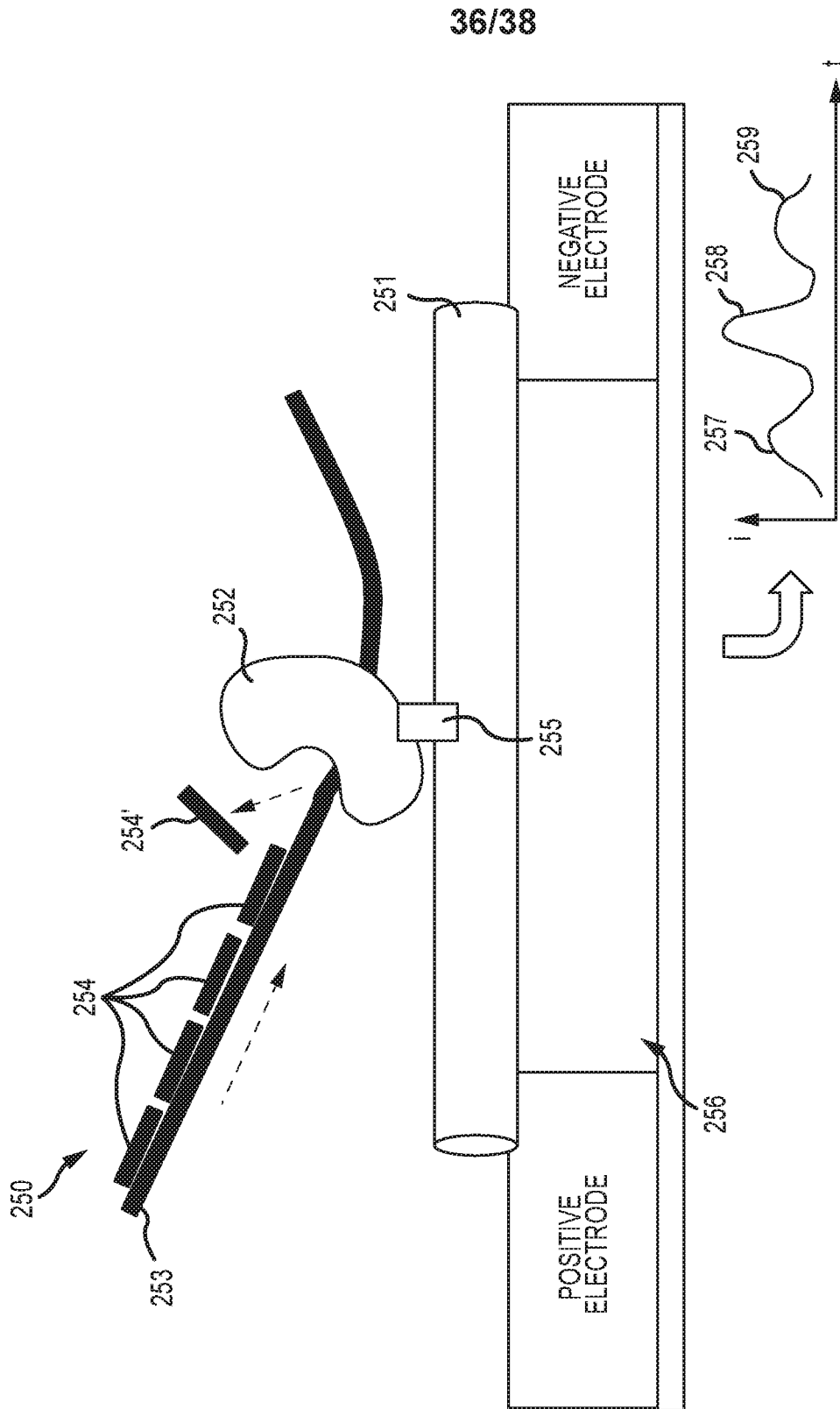


FIG.25

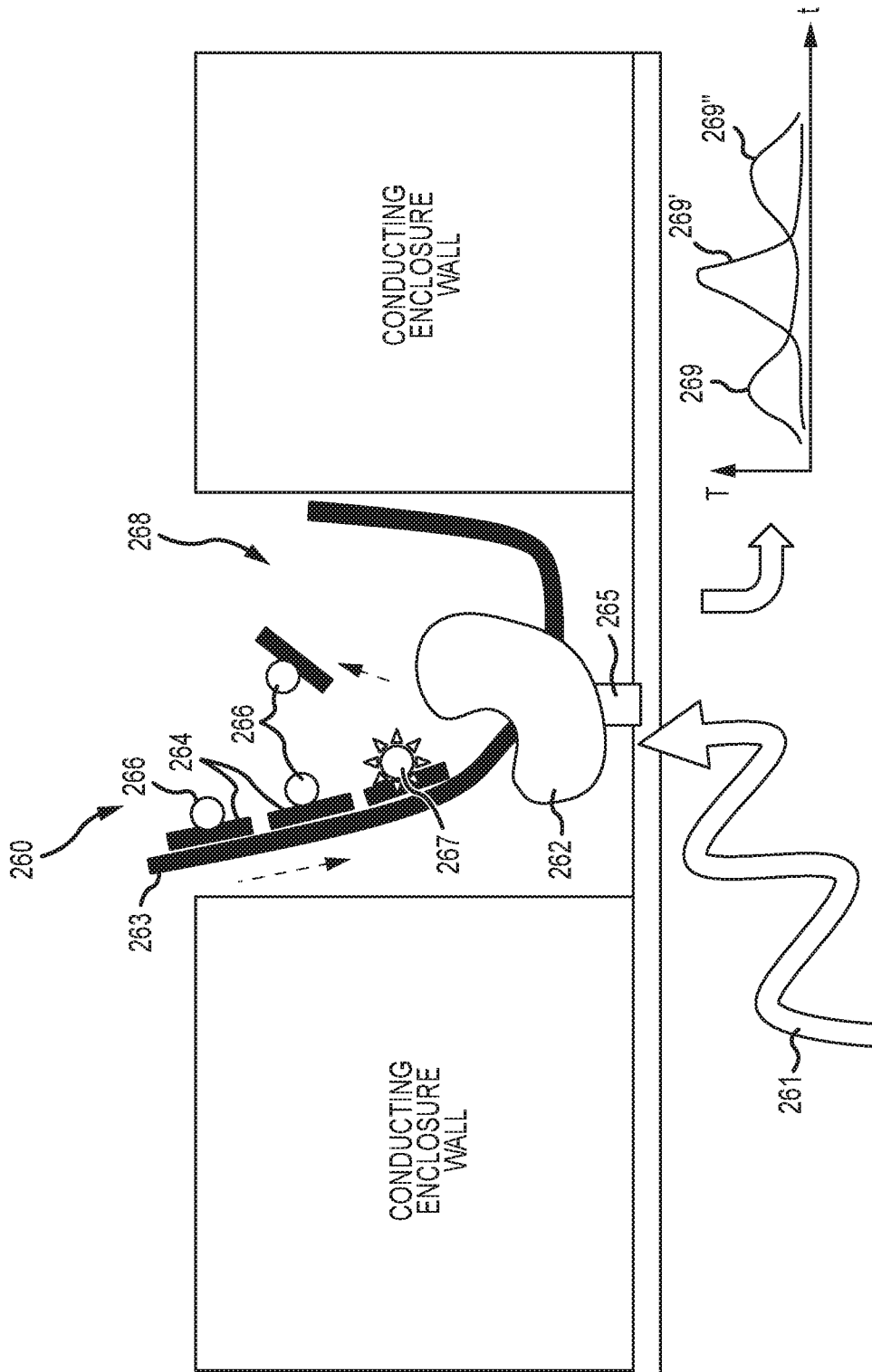


FIG.26

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2018/048873

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12Q 1/00; C12Q 1/02; G01N 27/26; G01N 27/30 (2018.01)

CPC - C12Q 1/025; G01N 27/3274; G01N 27/4163; G01N 33/48; G01N 33/49 (2018.08)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

USPC - 204/403.02; 205/782; 435/6.12; 436/501 (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2017/132586 A1 (ROSWELL BIOTECHNOLOGIES, INC.) 03 August 2017 (03.08.2017) entire document	1-40
P, X	WO 2018/132457 A1 (ROSWELL BIOTECHNOLOGIES, INC.) 19 July 2018 (19.07.2018) entire document	1-40
A	US 2017/0184542 A1 (CILAG GMBH INTERNATIONAL) 29 June 2017 (29.06.2017) entire document	1-40
A	US 2017/0234825 A1 (INTEL CORPORATION) 17 August 2017 (17.08.2017) entire document	1-40
A	WO 2012/152056 A1 (THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY et al) 15 November 2012 (15.11.2012) entire document	1-40

 Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

12 October 2018

Date of mailing of the international search report

09 NOV 2018

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774