I. INTRODUCTION

A. From a Single Neuron to Neural Ensembles

The daunting task of unraveling the physiological mechanisms that account for the operation of the human brain, a highly complex and self-adaptive biological system, formed by ~100 billion interconnected neurons (37), has become the true holy grail of neuroscience since the first images of brain circuits were produced, more than 100 years ago, by the skillful hands of Santiago Ramon y Cajal (256, 658). By the time Cajal’s histological approach was complemented by its electrophysiological counterpart, Sir Edgar Adrian’s metal microelectrode approach to record the electrical pulses produced by individual neurons (2), the neuron doctrine (732) had become established as the fundamental theory in the emergent field of brain research. This doctrine purports that individual neurons work as the functional unit of the brain through processing and transmission of electrophysiological signals.

Despite the undeniable success of the neuron doctrine, since the origins of modern neuroscience, researchers entertained physiological models of brain function in which populations of neurons performed the fundamental job of generating functions and behaviors. Indeed, the Italian anatomist, Camillo Golgi, who shared the 1906 Nobel Prize in Medicine and Physiology with Ramon y Cajal, was the first to introduce the term neural network, as a way to describe the underlying “functional module” of brain operation proposed in his reticular theory (308). According to Golgi’s view, brain tissue should work pretty much like the heart.
myocardium, forming a syncytium, or a continuous network of fused or tightly connected neurons (308). Cajal’s histological demonstration of the existence of synaptic clefts between neurons in most of the brain debunked Golgi’s view from an anatomical point of view and, as a consequence, the reticular theory was abandoned and the term network fell out of favor with most neuroscientists. Ironically, many decades later, clusters of neurons tightly linked via gap junctions (155) were identified in some key structures of the mammalian brain, including the inferior olive (506), hippocampus (443), and neocortex (486, 747). Because of the abundant existence of tight junctions, these clusters exchange information through electrotonic coupling. As such, they clearly typify the type of networks envisioned by Golgi.

During the century of intense work that followed the pioneering discoveries of Ramon y Cajal and Golgi, many other histological, electrophysiological, and imaging methods have been incorporated in the technical arsenal employed by neuroscientists to probe brain function at different levels of spatial and temporal resolution (FIGURE 1). For most of this continuous procession of new technological developments, the neuron doctrine continued to flourish. As a result, for the vast majority of the neuroscientific community, the single neuron remained the central focus of systems neuroscience for most of the 20th century (583).

Yet, since the late 1940s, a neural network-based view of brain function began to reemerge. Inspired by the pioneering work of Thomas Young on color coding in the early 19th century (881) and that of Charles Sherrington on spinal reflexes at the beginning of the 20th (733), theoreticians, such as Donald Hebb (352), and neurophysiologists, such as John Lilly (499), proposed that the true functional unit of complex brains, such as ours and those of other mammals, is represented, according to Hebb’s own terms, by “...a diffuse structure comprising cells in the cortex and diencephalon, capable of acting briefly as a closed system, delivering facilitation to other such systems (352).”

Even though Hebb’s masterpiece work, *The Organization of Behavior* (352), published in 1949, launched the modern era of neural population coding in systems neuroscience, few took notice at the time of its publication. By then, neurophysiologists were primarily engaged in characterizing the physiological properties of individual neurons, either using the classical Adrian’s approach of extracellular single neuron recording (2, 372) or a new methodological breakthrough of that time: intracellular single neuron recordings using sharp glass electrodes (92, 503, 781). Certainly, the multiple technological challenges involved in developing techniques for recording simultaneously from large populations of individual brain cells, even in anesthetized animals, let alone awake preparations, kept most experimentalists away from trying to test experimentally the

---

**FIGURE 1.** Temporal and spatial resolution for different techniques to study the brain and interact with its circuitry. **A,** techniques for recording and visualizing. **B,** electrode techniques for recording and stimulation. **C,** techniques for stimulation and making lesions. (Adapted with permission from Sejnowski TJ, Churchland PS, Movshon JA. Putting big data to good use in neuroscience. *Nature Neurosci* 17: 1440–1441, 2014. Reprinted by permission from Macmillan Publishers Ltd.)
ideas about neural populations introduced by Hebb. Thus, if one combines these notorious and, at the time, insurmountable technical difficulties, with the stupendous success and widespread acceptance of Cajal’s neuron doctrine, amplified significantly in the 1960s and 1970s by the enormous impact of the Hubel and Wiesel’s characterization of single neuron physiological properties in the visual cortex, it is no surprise that a neural population view of brain function had to wait for four long decades before it could begin receiving serious experimental attention by neurophysiologists.

Among the experimental studies that rekindled the interest in neural populations in neurophysiology during the 1980s was the pioneering work of Apostolos Georgopoulos on directional coding in the primate motor cortex (292, 295, 299) and John O’Keefe’s discovery of place neurons in the rat hippocampus (600). Georgopoulos’ findings that individual neurons in the primate primary motor cortex are broadly tuned to the direction of arm movement (294, 295, 423, 710) and that populations of such neurons, rather than an individual M1 neuron, have to be pooled together to compute the direction in which a monkey is about to move its arm (292, 299) brought to the forefront the much neglected, if not forgotten, Hebbian’s view of the neural population basis of brain function. By the mid-1990s, the introduction of new electrophysiological methods for chronic multielectrode recordings in freely behaving animals triggered a new phase of neural ensemble physiology (581, 582, 584, 857, 858). In this experimental approach, arrays or bundles of microelectrodes, originally made of fine insulated metal filaments, were chronically implanted across multiple cortical and subcortical structures of the brain of rodents and primates and yielded viable single and multiunit activity for long periods of time, which today can reach several years for the same animal (711) (FIGURE 2). By the early 1990s, this new approach was yielding simultaneous recordings of ~12–24 neurons in freely behaving rats. These recordings lasted for several weeks or even months (578, 579, 584). By mid-1995 the yield increased to ~50 simultaneously recorded neurons, with the added capability that individual neurons could be recorded in up to five different subcortical and cortical structures that defined a given neural pathway (i.e., the rat trigeminal somatosensory system) (578). By 1999–2000, this recording benchmark reached 100 neurons, and such simultaneous recordings could be obtained in both awake rats and monkeys (580–582). At that time the modern concept of BMIs was proposed by John Chapin’s and Miguel Nicolelis’ laboratories working together (124, 575, 852).

**B. Emergence of BMIs**

The modern era of BMIs emerged precisely at the time the methods for chronic multi-electrode recordings were consolidated in rodents and started to make their transition to primates, first to New World monkeys (582, 852) and then to rhesus monkeys (114, 463, 580). Around that time, the term BMI was introduced for the first time in the systems neuroscience literature (575). The main goal of these studies was to investigate physiological properties, including the ability of neural ensembles in the sensorimotor cortex to encode information and express plastic adaptations while freely behaving animals learned new motor tasks (460). However, with the publication of a series of original studies, conducted in rats and monkeys in the early 2000s, it soon became apparent that BMIs could also serve as the foundation of a new generation of neuroprosthetic devices aimed at restoring mobility to patients severely paralyzed due to trauma to the nervous systems, notably spinal cord injuries (SCIs) or neurodegenerative diseases. FIGURE 3 shows the original schematic description of this idea, proposed in the early 2000s, as an envisioned direct link between a human brain and a robotic arm. FIGURES 4 AND 5 illustrate how the original BMI control scheme was adapted to become the experimental paradigm employed with rhesus monkeys. For the most part, until today, most laboratories around the world that work with upper limb BMIs
continue to employ the elements depicted in FIGURES 3–5, which were originally published about 17 years ago (124, 852). Indeed, most researchers working in the field would consider FIGURE 3 as the standard or operational definition of a BMI, which can be described as a reciprocal link between an animal or human brain and an artificial actuator, such as a virtual or robotic arm or leg, which allows the subject to utilize its own volitional electrical brain activity to control the movements of such an actuator, while receiving continuous feedback information from it.

As we will see throughout this review, many aspects of this standard definition have been expanded significantly over the past two decades. For example, as illustrated in FIGURE 5A, large variety of brain signals, ranging from single units (114, 124, 466, 583, 852), to local field potentials (LFPs) (264, 670, 750), electrocorticography (ECoG) (483, 484, 697, 759, 837, 839, 856), electroencephalography (EEG) (7, 71, 141, 169, 358, 412, 627, 865), all the way to magnetic resonance imaging (MRI) signals (666, 745, 755, 847, 879), have been utilized as the source of voluntary motor activity needed to control an artificial actuator. By the same token, BMIs have been now designed to control a large variety of actuators, including computer cursors (114, 283, 463, 692, 725, 794, 864), digital communication systems (10, 71, 101, 247, 545, 559, 672), robotic limbs (114, 152, 360, 463, 795, 817), robotic exoskeletons (156, 456, 467, 836), avatar bodies (151, 377, 435, 598, 657, 835), drones (426, 457), and wheelchairs (169, 511, 553, 636, 890). BMIs have also been coupled with a variety of traditional medical devices (522). For instance, recently a BMI core was used to allow patients to control stimulators that activate their own muscles through functional electrical stimulation (FES) (83, 633) (FIGURE 6).

To extract the type of information needed to control such a vast list of actuators, a multitude of mathematical and computational approaches have been proposed as potential real-time decoders of voluntary motor activity generated by the brains of animals and human subjects (49, 449, 490).
Moreover, a variety of signals and strategies for delivering sensory feedback to subjects operating a BMI have been proposed, including visual feedback through computer screens (114, 377, 692), tactile and proprioceptive cues via haptic displays (128, 597), and even direct intracortical microstimulation (ICMS) applied to the primary somatosensory cortex (S1) (262, 597, 598).

To give an idea of how energized the field has become, even BMIs that involve the collaboration of multiple animal brains have been demonstrated in rat and monkeys (614, 616, 657). Notwithstanding these countless innovations, FIGURES 3 AND 4 still are useful to describe the basic elements that constitute a BMI. These include the use of sensors (e.g., multielectrode arrays) to sample large-scale brain activity; multi-channel electronics that amplify, filter, and digitize these signals; a computational engine responsible for real-time extraction of motor commands from the raw brain activity; an artificial actuator that performs motor tasks; and a stream of feedback signals delivered to the subject’s brain.

Merely 17 years after the modern age of BMI research was launched, the tremendous impact of this paradigm in neuroscience can be measured by a variety of metrics. For example, a bibliographic search in Google Scholar using the terms brain-machine interface and a closely related expression, brain-computer interface, yields more than 40,000 publications during the past decade alone. Meanwhile, several of the key experimental and clinical papers in the field have surpassed 1,000 citations. Several prominent books have been published on different aspects of BMIs (61, 106, 106).
During the same period, an estimated $800 million United States dollars have been invested in BMI research worldwide. Since the introduction of experimental BMIs in the late 1990s, many applications have emerged for healthy subjects, outside the domains of basic and clinical research. BMIs for computer gaming (6, 9, 257, 342, 489, 493, 531, 532, 775, 874), EEG-based that detect drowsiness in drivers (137, 286, 501, 504, 636, 685, 810), and even BMIs for education (371, 526) are just a few examples of this parallel line of BMI development outside biomedical research. As a result, many started to consider BMIs as a method to augment human neural and physiological functions, such as cognitive abilities (242, 518, 519, 539, 888) and motor performance (148). Despite being very interesting subjects for debate, none of these latter areas of BMI application will be covered here.

In this review, we start with a brief history of BMIs, followed by a description of major classes of BMIs. Next, we spell out the major components for building a BMI: sensors to record large-scale brain activity, decoding algorithms for extracting behavioral variables from brain signals, the means to deliver sensory feedback to the brain, and external actuators controlled through BMIs. Finally, we will discuss BMI applications to restore mobility and, eventually trigger partial neurological recovery in severely paralyzed patients.

II. HISTORY OF BMI RESEARCH

A. Early Studies

As mentioned in the introduction, the history of BMIs is intimately related to the effort of developing new electrophysiological methods to record the extracellular electrical activity of large neuronal populations using multi-electrode configurations. Such an essential component of modern BMI architecture was pioneered in the 1950s, by John Cunningham Lilly, then a principal investigator at the National Institutes of Health. In his experiments, Lilly was able to implant 25–610 electrodes, either on the pial surface of the cortex or intracortically, in adult rhesus monkeys (498) (FIGURE 7). In addition to recording field potentials (25 channels at a time) with these electrodes while animals exhibited a variety of behaviors and states (arm movements, sleep, etc.), Lilly also applied electrical current through those electrodes and elicited movements in both anesthetized and awake monkeys (496). He observed that motor responses could be evoked from many cortical sites, including M1 and S1. Lilly concluded that there was no clear-cut separation between cortical regions presumed to be motor alone or sensory alone. He suggested that these areas be named sensorimotor instead (497).

The next intermediary step in the development of the BMI concept can be traced back to the introduction of “EEG biofeedback” or “neurofeedback,” which became very popular in the 1960s and 1970s, in a variety of experimental settings (179, 258, 405, 407, 622, 749, 764, 777, 843, 867). In these studies, subjects were provided with an indicator of their own neural activity, for example, auditory or visual feedback derived from EEG recordings, which assisted them with self-regulating those neural signals. David Nowlis and Joe Kamiya (405, 595, 596) recorded EEGs in animals and human subjects, and converted them into sound. Aided by this type of neurofeedback, subjects gained some level of volitional control over their own EEG activity. Maurice Sterman and his colleagues converted EEGs of epileptic patients into lights and tones, and achieved seizure reduction with this type of neurofeedback training (763–766).

According to Daniel Dennett (193), the first experiment in which human subjects sent brain-derived signals to command an external device was described in 1963 by Grey Walter in a talk to the Osler Society at Oxford University. Since Walter himself did not publish this presentation, Dennett’s anecdote may not be completely accurate or verified.
Published literature, however, confirms that Walter and his colleagues implanted multiple electrodes (up to 1,000) in the cortex of neurological patients and used these implants for monitoring cortical field potentials over a period of several months (831). In the experiments described by Dennett, Walter recorded motor cortex readiness potentials preceding movements, also called Bereitschaftspotentials, while his patients periodically pressed a button to advance slides in a slide projector. The button presses were self-paced. The readiness potentials led the movement by approximately half a second and were sufficiently strong to be detected by Walter’s recording equipment. Next, Walter succeeded in creating a direct link between each patient’s motor cortex and the projector. In this experimental condition that we would now call brain control, the button was electrically disconnected from the projector and the slides were advanced by motor cortical readiness potentials. It came as a surprise to the patients that the projector responded to their will even before they physically initiated the movement. Although Walter’s experiments can be considered as the first proof of concept of the possibility of building BMIs, he never published these results or interpreted them in the context of BMIs, even though in his earlier career in the 1950s, he conducted research on robots with artificial brains (829, 830).

By the end of the 1960s, researchers at the NIH Laboratory of Neural Control started to experiment with the possibility of utilizing recordings from cortical neurons to control artificial actuators (270). They were also interested in using direct connections between brains and external devices to restore hearing to the deaf, walking to the paralyzed, and vision to the blind (701). This NIH-led research was conducted with some universities and medical schools participating as subcontractors. Karl Frank, the NIH laboratory head, proclaimed, “We will be engaged in the development of principles and techniques by which information from the nervous system can be used to control external devices such as prosthetic devices, communications equipment, teleoperators . . . and ultimately perhaps even computers” (269).

In their initial study, the NIH team implanted five micro-electrodes in the primary motor cortex (M1) of rhesus monkeys and then recorded action potentials generated by 3–8 M1 neurons, while animals performed a motor task that required them to flex and extend their wrists (374). Since the eventual goal was to convert these neuronal signals into the movements of an external device, the researchers probed whether wrist movements could be predicted from the recorded activity of small neuronal populations. They utilized multiple linear regression as a prediction algorithm. The algorithm took neuronal rates as inputs and returned movement kinematics as the output. A decade of this research eventually resulted in the demonstration of a real-time neural control (702): a rhesus monkey with 12 micro-electrodes implanted in M1 for 37 mo learned to move a cursor on an LED display using its own neural activity as a direct source of motor commands.

During the late 1960s, Eberhard Fetz and his colleagues conducted experiments in which they utilized the electrical activity of single neurons recorded in monkey M1 as a source of neurofeedback (254). Using this apparatus, monkeys learned to self-regulate their own single-neuron activity. Typically, one neuron was tested at a time. The neuronal electrical firing rate was converted into either auditory (a click for each spike) or visual (deflections of an arrow meter placed in front of the animal) feedback. Monkeys learned to volitionally modulate the activity of each individual M1 neuron to reach a particular level of firing required to obtain a reward (FIGURE 8). While Fetz emphasized the neurofeedback aspect of such operant conditioning experiments, Brindley and Craggs (89, 168) employed epidural recordings of motor cortical field potentials in the frequency band 80 to 250 Hz in baboons to test the possi-
bility of creating a motor neuroprosthesis that recognized specific movements produced with the arms or legs. Moreover, Craggs (167) used baboons with complete spinal cord transections at the midthoracic level as a model of human paraplegia, while he recorded motor commands directly from the cortical representation of the foot disconnected from its spinal cord projection area.

At the same time these laboratories experimented with extraction of motor signals from the brain and/or using them to generate neurofeedback, another line of research focused on ways to deliver information to the brain using electrical stimulation, either applied to peripheral nerves (153, 354) or to the central nervous system (90, 91, 495). This work led to early attempts to build sensory BMIs that strived to restore normal perception to patients suffering with neurological conditions that induced significant sensory deficits. From these pioneering studies, the work on cochlear implants (Figure 9) eventually reached the most spectacular results (198, 218, 219, 367, 510, 743, 855). In parallel, some progress was achieved in the development of visual cortical prostheses pioneered by the groups led by Brindley (88, 90, 91) and Dobelle (200–203). These researchers applied electrical current to the visual cortex of blind patients through grids of surface electrodes. Using this apparatus, blind subjects could perceive light spots, phosphens, and learned to recognize simple visual objects composed of several phosphens.

Also in the 1960s, Bach-y-Rita and his colleagues started to develop visual substitution systems for the blind, based on tactile stimulation of the skin on the patient’s back (39, 40). This technique became known as vision substitution by tactile image projection. The apparatus employed, called a haptic display, consisted of 400 solenoid stimulators arranged in a 20 by 20 array. The tactile stimulation was applied to the surface of the patient’s back and attempted to reproduce, through the sense of touch, visual images captured by a video camera (Figure 10). After being trained for 10 h, blind patients learned to recognize objects and their positional relationship in a room, as well as landmarks, such as the room’s door frame (40).

**B. Explosive Development in the Late 1990s**

Despite the initial push observed in the 1960s and 1970s, research on direct links between brain and machines experienced a decline during the next 20 years. Clearly, the lack of major technological breakthroughs in the area of multi-channel neuronal recordings during that period prevented the field from taking off and fulfilling the auspicious poten-
tial envisioned by the early proponents of this research program.

Around the mid-1990s, however, the required technological innovation materialized with the introduction of a new multi-electrode design that allowed bundles or planar arrays to be built using flexible and insulated metal filaments, known as microwires (108, 578, 584, 585). John K. Chapin, one of the pioneers of microwire implants, discovered that if a blunt tip was left exposed in an otherwise insulated metal microwire, made of stainless steel, individual single-units could be recorded for many weeks or even months from the forepaw representation area of M1 and S1 of freely moving rats (125, 734). A few years later, Miguel Nicolelis and Chapin chronically implanted multiple bundles and/or arrays of Teflon-coated stainless steel microwires in the ventral posterior medial nucleus (VPM) of the thalamus of adult rats and recorded the simultaneous electrical signals produced by up to 24 of these thalamic neurons in awake and freely moving rats (579, 584). A year later the same technique was validated in primates, allowing the Nicolelis Lab at Duke University to record from multiple cortical areas in awake owl monkeys (582).

Since the recording properties of chronic microwire implants lasted for several months in both rats and owl monkeys, Chapin and Nicolelis found this technique suitable for testing the concept of linking the brains of rats and monkeys to artificial actuators and investigating whether these animals could learn to control external devices using only their brain electrical activity. In 1999, Chapin and Nicolelis published their first BMI study where rats learned to use the combined extracellular activity of up to 46 neurons distributed across multiple subcortical and cortical relays of the rat trigeminal somatosensory system, in awake, freely behaving rats. These experiments marked the first time simultaneous neuronal population activity, originating from multiple processing levels of a mammalian sensory system, was measured in the same animal subject. Three years later the same technique was validated in primates, allowing the Nicolelis Lab at Duke University to record from multiple cortical areas in awake owl monkeys (582).
Curiously, also in 1999, without knowledge of the work carried out by Chapin and Nicolelis in animals, a group led by Niels Birbaumer at the University of Tubingen in Germany pioneered their version of a direct link between a brain and a computer in locked-in patients (71). Birbaumer chose to name this paradigm brain-computer interface (BCI), a term that had been introduced in the literature by Jacques Vidal in 1973 (823). Birbaumer’s BCI allowed locked-in patients to communicate with the external world using slow cortical potentials recorded via a noninvasive technique, EEG recordings, to control computer software for spelling. Using this system, locked-in patients became capable of writing messages on the computer.

Overall, the original papers by Chapin, Nicolelis, and Birbaumer mark the beginning of the modern age of research on BMIs. In the case of intracranial BMIs, the focus of this review, advances in multi-electrode recording methods, combined with the introduction of faster digital computers running new computational algorithms for extracting motor signals from brain-derived signals, triggered a phase of very fast growth in the field. Thus, following the original demonstrations in rats and New World monkeys, the next important milestone of the field was the translation of the BMI paradigm to rhesus monkeys, the conventional experimental animal model for exploring neurophysiology of advanced motor behaviors and cognition. In a quick succession, three different groups published their results in this primate species (114, 725, 794). In a span of 12 mo, the BMI paradigm in rhesus monkeys incorporated the use of a series of novel actuators, such as a computer cursor (725, 794), and a robot arm capable of producing both arm reaching movements and hand grasping (114). With this latter addition, the Nicolelis laboratory demonstrated that the same pool of recorded cortical neurons could be employed to simultaneously extract hand gripping force and arm position and velocity from multiple frontal and parietal cortical areas in awake rhesus monkeys (114, 463).

In 2004, the Nicolelis laboratory also reported the first demonstration that ensembles of subcortical neurons, recorded intraoperatively with microwire bundles, could be employed to extract hand movements in awake and conscious human subjects (620). These recordings were obtained during a neurosurgical procedure in which Parkinsonian patients received a deep brain stimulator. During a brief intraoperative period of 10–15 min, up to 50 neurons located in the subthalamic nucleus and thalamic motor nuclei were recorded while these patients played a one-dimensional video game by exerting a gripping force with one hand. This study revealed that the same computational algorithm, multi-linear regression, employed by Wessberg et al. (852) and Carmena et al. (114) in owl and rhesus monkeys, respectively, could be used to extract hand movement patterns from human subcortical signals.
A couple of years later, in 2006, the group led by John Donoghue reported the operation of a BMI in one patient chronically implanted with a different multichannel recording technology approved for human trials (362). Their implant, named the Utah probe, was a $10 \times 10$ array of silicon-etched rigid needles (111, 537, 593). The Utah probes were inserted in the cortex ballistically using a pneumatic gun (680), the method adopted for this implant to avoid the “bed of nails effect,” where a slowly inserted dense electrode array produces cortical dimpling and trauma (76, 111, 668). A total of two patients were implanted with this device in 2006 (362), one of whom experienced implant malfunctions for the first 6 mo followed by 2 mo of recordings. Both patients used a BMI to control two-dimensional movements of a computer cursor.

As discussed below, recordings with Utah array suffer from biocompatibility issues (220, 251, 679), which in our opinion should preclude them for further use in human subjects. Shortly after implantation, this probe can produce a significant tissue lesion and, hence, become encapsulated by glia and protein deposits, as a result of the local inflammatory reaction. Usually, this process renders the Utah probe unusable for single-unit recordings after a few weeks/months (134). Groups that rely on this probe usually resort to the threshold-crossing method to detect useful neuronal activity (134, 803), a maneuver that discards well-established neuronal recording quality criteria (580) and increases the likelihood of recording noise instead of neuronal spikes, by mistakenly recording mechanical, electrical and EMG artifacts as if they represented valid neural activity. As such, this recording technology cannot be considered as the final solution for clinical BMI applications, even though currently this method is commonly used in human trials. Practical solutions will clearly require better recording reliability, stability, and longevity standards to become accepted by patients and clinicians.

To achieve better biocompatibility of a brain implant, in 1989, Philip Kennedy implanted an ALS patient with a neurotrophic electrode loaded with nerve growth factors that induced growth of nerve fibers into the electrode tip (418, 420, 422). The study reported that the patient learned to produce on/off neural control signals that were detected by the electrode. While the same group continues this research until now (98, 327), neurotrophic electrodes were not adopted by other groups and their effectiveness remains difficult to evaluate.

In addition to a few more clinical demonstrations of BMIs that controlled computer cursors (3, 731), upper limb robotic prostheses (152, 360) or an FES system for the hand (83) using either populations of M1 or other cortical neurons, several innovations were incorporated to the traditional BMI approach over the past several years. In 2009, the Nicolelis laboratory published the first BMI approach to decode kinematics of bipedal walking in rhesus monkeys (261) (FIGURE 12). Two years later, in 2011, the same laboratory implemented, for the first time, a method for multi-channel ICMS as a tool to deliver direct tactile feedback to the subject’s somatosensory cortex in a BMI setup (598) (FIGURE 13). This new
paradigm was named a brain-machine-brain interface (BMBI). In those experiments, rhesus monkeys performed an active tactile discrimination task using a BMBI that both generated motor commands and delivered artificial tactile feedback directly to the animal’s brain. To perform the exploration task, monkeys had to control the movements of a virtual hand, using their motor cortical activity alone, to scan the surfaces of up to three visually identical virtual objects shown on a computer screen. The position of the virtual objects changed with every new trial. As the virtual hand touched each virtual object, temporally patterned ICMS was applied to the animal’s primary somatosensory cortex to mimic the virtual textures. To receive a juice reward, the monkey had to find an object associated with a particular virtual texture, and hold the virtual hand over that object. After a few weeks of training, two monkeys learned to perform this task at levels similar to those attained when the control signal to the virtual arm came from the joystick that the animals moved with their own biological hands.

In parallel with these developments, considerable efforts have been made by many laboratories to design better real-time decoding algorithms. Krishna Shenoy and his colleagues (692) reported a “high-performance BMI” that relied on the strategy of flashing potential targets in a rapid succession on a computer screen. They recorded from small neuronal populations in dorsal premotor cortex and found that the firing of these neurons reflected the target location. Target locations could be decoded from these neuronal firing modulations using recording intervals as short as 250 ms, which allowed the BMI to reach an information transfer rate of 6.5 bits/s. The authors argued that the observed neuronal responses represented the monkeys’ motor preparatory activity for arm reaching movements rather than merely visual responses to the targets, and therefore could be useful for controlling a motor BMI that automatically directs the cursor to the target once its location is determined. The same group developed an improved algorithm for continuous cursor control (307). The improvement was achieved using the “recalibrated feedback intention-trained Kalman filter” that was trained using both the cursor position in screen coordinates and an estimate of intended velocity based on the relative location of the cursor and target.

Moving in the same direction, our laboratory achieved considerable improvement of real-time decoding by employing an unscented Kalman filter that used position, velocity, and speed as state variables and incorporated nonlinear relationships between the neuronal rates and these variables (491). More recently, Jose Carmena and his colleagues (728, 729) reported an improved temporal resolution for an adaptive decoder that modeled spikes as a point process. Overall, these efforts resulted in a large variety of BMI decoding algorithms from which designers can choose depending on the requirements for their experimental or future clinical implementations.
clinical practice by Niels Birbaumer (71), have undergone a considerable expansion (14, 74, 246, 358, 815, 863). BMIs for humans are often referred to as BCIs, including both invasive and noninvasive systems. It is worth noting that the first publication on a human controlling a robot with EEG activity dates back to 1988 (85). In this report, subjects issued start and stop commands to a robot by closing and opening their eyes, the well-known procedure to activate and deactivate alpha waves (60, 587). Recently, a motor imagery-based BCI was used by the Walk Again Project, an international nonprofit research consortium, to allow complete paraplegic patients to use EEG to control the start and stop sessions of bipedal walking of a lower limb robotic exoskeleton (737). Overall, BCI research yielded many practical applications, such as BCIs that use EEG signals to control computer cursors (240, 864, 865), computer-assisted spellers (102, 245, 247, 421, 545, 832), wheelchairs (121, 411, 511, 553, 811, 833), and exoskeletons that restore bipedal walking (156, 836).

III. BMI CLASSIFICATION

A. Classification by Function

Several BMI classification schemes have been proposed heretofore. One scheme classifies BMIs according to the physiological function they are intended to emulate. Here, BMI systems are commonly categorized as follows: 1) motor, 2) sensory, 3) sensorimotor (or bidirectional), and 4) cognitive. The recent introduction by our laboratory of BMIs that incorporate multiple brains of different subjects (614, 616, 657) adds one more BMI class, which we named Brainet (657).

Motor BMIs reproduce motor functions, such as upper (114, 152, 817, 852) and lower limb (261) movements or whole body navigation (656, 873). Sensory BMIs aim at reproducing sensations, while sensorimotor BMIs combine the motor and sensory components in a single application (57, 597, 598). Cognitive BMIs (19) enable higher-order
brain functions, such as memory (62), attention (279, 512), and decision-making (343, 564). Finally, Brainets involve the implementation of shared BMIs requiring the participation of multiple subjects since these systems require the combination of electrical activity of multiple brains simultaneously to operate properly (657). Although seemingly straightforward, this classification scheme follows the traditional labeling of brain areas as motor, sensory, or higher order (associative), a simplified description, which is fundamentally wrong in most cases. There is mounting evidence against such a parcellation scheme of brain functions and in favor of a more distributed mode of information processing in the primate cortex (96, 253, 497, 583, 675). Thus sensory and motor signals are typically multiplexed by cortical neurons, including the areas considered to be purely motor or purely sensory (470). Indeed, even the primary motor and somatosensory cortical areas have been shown to concurrently process both motor and somatosensory information (229, 277, 497, 500, 647), and also represent visual signals (738), reward amount (529), and even cognitive processes, such as mental rotation (297, 300) and encoding serial order of stimuli (115).

Based on our own BMI work, we have argued for more than a decade that highly distributed neuronal ensembles represent the true physiological unit of the nervous system (583), a proposition that is supported by an extensive literature (107, 109, 253, 302, 339, 477, 648, 686, 687). In a distributed neural circuitry, there are no exclusive functional specializations at the level of individual neurons. Rather, single neurons multiplex several functions, and the best decoding can be achieved by sampling large numbers of neurons from multiple brain areas, simultaneously. We have proposed, therefore, that as the BMI field advances, this new theoretical view of brain function will lead to almost universal acceptance of a distributed principle of neuronal activity sampling in BMI applications. Since our first BMI studies (852), we have consistently demonstrated that large-scale neuronal recordings from multiple cortical areas are imperative for building BMIs that are versatile, robust, efficient, and clinically relevant (466, 575, 583). Another step in this direction was made when the BMI paradigm was introduced (598). In a BMI, motor and sensory streams of information are handled simultaneously to facilitate sensorimotor processing. In the same context, the concept of Brainets takes multitasking BMI designs to the next level by combining multiple brains into a higher order computational entity (614, 616, 657).

B. Classification by the Level of Invasiveness

Since the inception of modern-era BMIs, two major approaches have dominated the field: intracranial or invasive (124, 852) and noninvasive (71) systems. Accordingly, it is common to classify BMIs by their level of invasiveness. This division is important for a variety of reasons, the major one being a safety concern. Invasive BMIs require a neurosurgery procedure that involves opening the scalp and skull and penetrating the brain tissue, albeit only for a few millimeters, for systems relying on cortical signals. These procedures carry a risk of tissue damage and/or infection, particularly if the implant is not fully contained within the body and has external parts, like wires connected to extracranial recording hardware, as was the case in recent clinical trials in humans (83, 152, 360, 362). Noninvasive BMIs, on the other hand, do not carry such risks and can be implemented rather easily. For example, in the case of EEG recordings, the electrodes are simply placed on the scalp surface (588) through an easy and safe procedure, particularly if dry sensors (136, 266, 287, 329, 785) are used.

Since patient safety is of paramount importance, noninvasive BMIs are currently the default choice for clinical applications. Yet, their recording quality is insufficient in many cases, causing EEG-based BMIs to be rather slow systems. Indeed, EEGs represent attenuated and filtered brain activity, which combines synchronous electrical signals, produced by many millions of neurons. Since these signals have to travel through bone and skin prior to reaching the scalp sensors, EEG signals lack fine spatial resolution and do not provide the kind of precise task-related neuronal signals that can be obtained from intracranial recordings (466). Similar limitations characterize all noninvasive recording methods that measure neuronal signals at a distance from their source.

In contrast, in invasive BMIs, recording sensors are brought close to the very source of generation of neural activity: the single neurons that code information through trains of action potentials (157, 318, 487). Current intracranial BMIs usually employ extracellular recording methods that allow one to sample and discriminate action potentials generated by hundreds of individual cortical neurons (466, 583, 711, 767). The more microelectrodes are implanted; the more neurons can be sampled simultaneously. Moreover, the same microelectrodes can also record local field potentials (LFPs), which represent combined potentials of large (on the order of tens of thousands) neuronal populations (192, 289, 373, 398, 571). Additionally, the same implanted microelectrodes that are used for recordings can also be used for the delivery of electrical microstimulation that, depending on the stimulated area, can influence sensory, motor or cognitive processing (41, 45, 48, 146, 704, 705, 719, 770).

An intermediary approach, known as electrocorticography (ECoG) can be considered as a semi-invasive method since it requires a craniotomy but does not involve sensors penetrating the nervous tissue. ECoGs are recorded with a grid of electrodes placed on the brain’s surface; dura mater may be left intact (i.e., epidural ECoG) or open to allow closer contact between the electrodes and the cortex (subdural ECoG) (24, 94, 357, 530, 663, 696, 828). ECoG recordings...
have better spatial and temporal resolution than EEG, but they cannot be used to reliably detect single-neuron spikes. Chronic ECoG recordings can last many years in animals (103, 123, 131, 508, 682) and humans (845, 870). Overall, ECoG has many advantages compared with EEG, and it is not as invasive as penetrating implants. Still, it remains controversial whether or not ECoG-based BMIs can rival single units in terms of BMI performance and accuracy. A variety of intracranial and invasive approaches are reviewed below in the section on neural recording methods.

C. Classification by the Origin of Neural Signal

Since the late 1990s when the modern BMI concept and design was introduced, the majority of BMIs have utilized neural signals recorded from cortical areas of animals or human subjects (71, 114, 124, 152, 362, 377, 794, 852, 864). Such an abundance of cortical BMIs is not surprising because the cortex is the largest and most advanced brain structure, which is also the easiest to access with the recording sensors. Among cortical areas utilized in BMIs, M1 is the most commonly implanted area (114, 725, 852) because neuronal discharges in M1 are clearly correlated with different movement parameters (227, 294, 554). Additionally, recordings from premotor cortex are employed in BMIs as a source of motor commands to control ongoing movements (114) and preparatory signals that reflect motor planning before movements have started (692).

Our laboratory has long advocated recording simultaneously from multiple cortical areas as an efficient way to increase both the amount of information processed by a BMI and its performance and versatility (114, 377, 466, 583). We have routinely recorded from four to eight frontal and parietal cortical areas in rhesus monkeys to operate a variety of BMIs and observed that any of those cortical areas provide useful information. The best BMI performance was usually achieved when neuronal signals from multiple frontal and parietal cortical areas were combined.

Lately, interest has increased to BMI systems that utilize subcortical recordings. These systems, in principle, could capture neural processing in cortico-subcortical loops (11) related to motor control (188, 189, 677), sensory processing (44, 396, 397, 447, 528), motivation (706–709), and skill learning (72, 212, 356, 689). It is noteworthy that subcortical recordings from the ventrolateral thalamus (VL) were utilized in the pioneering BMI study by Chapin et al. in 1999 (124) where VL neurons contributed to the BMI control of lever movements. Additionally, Patil et al. (620) relied on subcortical recordings from the subthalamic nucleus (STN) and ventral intermediate/ventral oralis posterior motor thalamus (Vim/Vop) to demonstrate, for the first time, that human patients could utilize real-time algorithms previously tested in monkeys to generate one-dimensional movements of a computer cursor (336). More recently, Koralek et al. (442) employed cortical and striatal recordings in awake, behaving rats to construct a BMI for abstract skill learning. In these experiments, rats learned to control the pitch of an auditory stimulus through a BMI. This learning was accompanied by increases in neuronal firing in the striatum. Additionally, stronger correlations developed between the cortical and striatal neurons. These findings further supported the claim, made in early BMI studies (114, 124, 463), that brain plasticity, in this case, corticostratal plasticity, plays a major role in learning to control a BMI. This is an important conclusion for the development of BMIs, since it means that severely disabled patients may be able to learn novel motor skills, through BMI training.

In the future, subcortical BMIs could contribute to treatment of neural conditions caused by disorders of subcortical processing, such as Parkinson’s disease. As a step in this direction, we analyzed pathological signs in neuronal populations in Vim/Vop and STN recorded in Parkinsonian patients (336), while they controlled a computer cursor by opening and closing their hands. Their task was to point to screen targets with the cursor. Vim/Vop and STN neuronal populations responded to target onset, and hand movements, as well as being correlated with hand tremor. BMI decoders extracted movement kinematics from the STN population activity even when those populations exhibited tremor-related oscillations. These findings indicate that, in the future, BMIs based on subcortical recordings could be used for monitoring signs of neurological diseases, evaluating medical treatments and even delivering real-time rehabilitation therapies, via implanted devices, without the need for continuous supervision.

D. Classification by BMI Design

Development during the last two decades resulted in several well-established BMI designs. Two broad classes of BMIs are represented by the so-called independent (endogenous) and dependent (exogenous) systems. Although this terminology is usually applied to noninvasive BMIs (i.e., BCIs), it is also applicable to intracranial BMIs. In an independent BMI, subjects self-initiate actions, for example, by imagining movements (604, 631, 651, 698) or even assisting themselves with overt movements of the limbs (114, 307, 463, 852). Such imagery and self-generated movements are controlled voluntarily by the subjects and, in principle, could be performed independently from any external stimuli (although some external stimuli are usually involved).

Dependent BMIs, as the name suggests, critically depend on the presence of an external stimulus and the triggered neural responses to this stimulus (246, 473, 717, 832). For example, a P300 EEG- or ECoG-based BCI monitors cortical responses to computer screen events and detects a stronger response to the stimulus attended by the subject (101, 209,
257, 545, 635, 805). A very similar design, mentioned above, was implemented in an intracranial BMI that evaluated premotor cortex responses to visual targets that rapidly flashed on the screen (692). The improvement in performance was achieved because the decoder received the information at the time of stimuli presentation and started to analyze neuronal data precisely after this event. Such utilization of the timing of external events is characteristic of all dependent BMIs. For example, BMIs that incorporate instructed delay tasks rely on a precise sequence of computer-generated task events for decoding (564, 731). While such supervised operation speeds up the decoding and makes it more reliable, dependent BMIs can operate only for a given set of rules, a property that clearly limits the user’s autonomy in choosing motor outputs.

In addition to independent and dependent BMIs, the term passive BMI (or passive BCI) has been recently introduced to describe a system that performs useful decoding of neural signals without considerable mental efforts of the subjects (885, 887). Passive BMIs could, for example, improve human interactions with a technical system by monitoring and decoding neural signals representing cognitive and emotional states, while making appropriate adjustments to the technical system.

Another criterion to classify BMIs is whether a subject performs overt movements while performing a motor task using the BMI. Early BMI demonstrations relied on overt movements to train the decoder and to operate a BMI (114, 261, 463, 852), whereas the next generation of BMIs excluded overt movements from both training and operation phases (377, 794). The goal of this latter modification was to mimic more properly the conditions of paralyzed patients who cannot produce overt movements. Withholding overt movements during BMI control dramatically changes neuronal ensemble activity patterns (463, 583), including a transient increase in correlation between cortical neurons, which tends to subside with further training (377).

Overall, having a monkey control a BMI without moving its own limbs is a much more difficult task for the animal compared with the BMI control assisted by overt behaviors. Some recent studies still allow monkeys to assist themselves in the BMI operation by producing overt limb movements (239, 307, 761). While these reports downplay the significance of this component (307), it is possible that the presence of overt movements contributed to the improvement in BMI decoding. We suggest that overt behaviors should be better monitored in BMI studies and compared with the changes in BMI performance.

In this context, we should also mention that the requirement to withhold limb movements, while producing actuator movements only through a BMI, defines a particular motor task by itself; one that resembles the well-known instructed delay task, where cortical neuronal firing modulations (143, 171, 535, 846, 859), and even spinal cord interneurons (653), change their activity in the absence of overt limb movements. Similar no-go requirements can be found in BMI task designs. For example, Ganguly et al. (284) attached monkeys’ arms to a KINARM apparatus and required the animals to maintain constant arm position on each trial, while moving a screen cursor using a BMI; any arm movements cancelled the trial. These experimental settings are virtually identical to a classical instructed delay task, with the exception that in the instructed delay task, the arm eventually moves to the target, whereas in the BMI task, arm movements were not required. Such a requirement to pay attention to the arm position is different from other studies where animals were not encouraged to pay attention to the arm position, which was irrelevant for the BMI performance (114, 377, 463). This difference is important because the results of such experiments are often interpreted in terms of incorporating an external effector into the brain’s own representation of the subject’s body (466). To validate such an interpretation, it matters whether the subject refocused attention to a new effector or continued to attend to the arm position and possibly used it as a reference for a BMI-controlled cursor.

In the early days of the field, the number of neurons needed for accurate BMI performance control used to be a controversial issue. During the early 2000s, research groups led by Donoghue (723, 725) and Schwartz (794) suggested that recording from just a few cortical neurons simultaneously could be sufficient for achieving good BMI control. In contrast, our laboratory has always argued that large neuronal ensembles are required to achieve optimal BMI performance (114, 261, 377, 466, 468, 583, 852). We reasoned that single neurons represent behavioral parameters of interest only partially and in a noisy way, and that combining contributions from many neurons both increases the information content and improves the signal to noise ratio of decoding. Additionally, it is easier to select neurons with the properties needed for decoding when there is a sufficiently large neuronal sample from which to select. As time passed, this dispute was resolved in favor of the large neuronal samples. Even the former proponents of small ensembles have now switched their approaches to embrace large neuronal samples as the only viable approach for clinically-relevant BMI applications (152, 360, 362).

IV. REPRESENTATION OF INFORMATION BY NEURONS AND THEIR ENSEMBLES

A. Tuning of Single Neurons

For the past six decades, an extraordinarily large neurophysiological literature has accumulated around the subject
of how ethologically meaningful information is encoded by individual neurons and neuronal ensembles (106, 142, 237, 293, 401, 578, 583, 671). Despite these extensive observations, we are still far from understanding the physiological mechanisms guiding the dynamical operation of neural circuits in mammals. This fundamental lack of knowledge, however, has not precluded BMI research from adopting an empirical approach, where parameters of interest are extracted from neuronal signals. Any decoding approach is always based on the existence of some degree of correlation between neural activity and those parameters of interest, but rarely on an unequivocal establishment of a causal relationship or a clear knowledge of the neural mechanisms involved. Yet, as we previously argued (466, 583), empirical approaches employed in BMI research could help us to uncover fundamental neurophysiological principles governing the operation of brain circuits.

The existence of a correlation between neuronal firing rate and a given behavioral variable is classically referred to as neuronal tuning. When one says that a neuron is tuned to a behaviorally relevant parameter, this simply means that the neuronal rate is correlated with that parameter, and this correlation is consistent. In this context, two key physiological properties of neuronal circuits have accounted for the feasibility of creating functional BMIs which can generate consistent motor outputs. First, BMIs have benefited from one of the early discoveries of neuronal ensemble physiology: that the trial-to-trial variability in firing rates of single neurons, which is often described as neuronal noise, can be significantly compensated by recording simultaneously from ensembles of neurons (253, 292, 462, 467, 583). In other words, combining contributions from many neurons reduces the uncorrelated noise produced by individual neurons while leaving intact the consistent component of firing modulations, best represented by the entire neuronal population. The second physiological property that proved to be indispensable for the proper operation of BMIs is the occurrence of neuronal plasticity, i.e., the ability of neurons to continuously adapt their tuning when exposed to novel task contingencies and external world statistics (see below).

The relevant literature on neuronal tuning starts with the work of Edward Evarts who pioneered the investigation of the physiological properties of single M1 neurons in the mid-1960s (233, 235, 236). By using sharp-tip electrodes, Evarts sequentially recorded extracellular electrical activity of single M1 neurons, while his monkeys remained awake and performed a variety of motor tasks (227, 228, 230–232, 234, 238). These now classical experiments revealed that M1 neurons were tuned to parameters such as muscle force and joint torque. For example, an M1 neuron would increase its firing rate when the monkey pulled a lever and decrease firing when the monkey pushed it. As such, M1 neuronal firing modulations coded the next push or pull movements performed in a task trial. Combining neuronal firing rates from many trials of each kind allowed Evarts to describe an average response curve, which was named peri-event time histogram (PETH).

Evarts’ original findings triggered a major push in systems neurophysiology, resulting in the widespread use of his technique for single-unit recording to characterize the tuning properties of individual neurons in various areas of the rhesus monkey’s brain. In the case of the motor cortex, the next breakthrough was produced by the seminal discovery of Apostolos Georgopoulos and his colleagues that M1 neurons exhibit broad tuning to the direction of arm movement (295, 299, 423, 710). To reach this conclusion, the Georgopoulos laboratory measured the firing discharge patterns of individual M1 neurons, while their monkeys performed arm reaching movements that started at an initial, central location and ended on peripheral targets (FIGURE 14). Analysis of these center-out movements revealed that the single neuron’s firing rate peaked for a particular movement direction, called preferred direction, and decreased gradually when the direction deflected from the preferred one. Georgopoulos graphically represented the relationship between neuronal rate and movement angle as the directional tuning curve (290, 710). He also proposed a cosine fit, where the neuronal rate was proportional to the cosine of the angular deviation from the preferred direction.

Further exploration conducted by many groups into the activities of cortical neurons during motor behaviors revealed many additional characteristics of neuronal tuning, such as representation of sensory signals (464, 505, 565), sensorimotor transformations (194, 297, 298, 400, 401, 860), simultaneous encoding of the motor goal and the direction of spatial attention (471), representation of multiple motor plans (143), and even encoding of cognitive variables by M1 neurons (291).

With the accumulation of these findings, it became progressively clear that the next step toward the understanding of brain encoding would require sampling the activity of populations of neurons recorded simultaneously, instead of recording one neuron at a time. In the case of BMIs, this step was imperative because single neurons or small neuronal samples (10–30 neurons) could not sustain BMI operational accuracy and stability.

B. Neuronal Ensemble Physiology

At present, it has become conventional for BMI studies to report that decoding of behavioral parameters from neuronal activity improves with an increase in neuronal sample (114, 261, 377, 463, 466, 468, 583, 852). The relationship between the sample size and decoding accuracy is depicted by a NDC (FIGURE 11), an analysis introduced by our lab-
oratory in 2000 (852). Following this original publication, this analysis quickly became a standard in the literature on neural decoding (50, 78, 283, 552, 564).

Based on a decade of BMI studies in our laboratory, we have proposed a series of principles of neural ensemble physiology, derived primarily from the analyses of NDCs...
Several of these principles can be immediately derived from the typical NDC, which shows a rapid rise in decoding accuracy for small neuronal samples followed by a slower rise for larger populations. Accordingly, the single-neuron insufficiency principle states that individual neurons carry low amounts of information. Conversely, the neuronal mass principle states that a given neuronal ensemble should reach a certain size, in terms of number of neuronal elements, for decoding accuracy to stabilize. At this point, decoding performance does not change substantially when a few neurons are added or removed. This stabilization of decoding accuracy should not be confused with saturation. Further improvement of decoding is possible, but a substantial increase in neuronal sample is needed to achieve an appreciable effect. The level of information grows as a function of the logarithm of the neuronal sample size. FIGURE 15 shows NDCs for several cortical areas and a variety of BMI experiments conducted in rhesus monkeys. It follows from the analysis of these NDCs that typically each cortical area contains some level of useful information, regarding the decoding of a given motor parameter. Accordingly, the distributed-coding principle postulates that information is represented by the cortex in a distributed way. This means that, for a given behavioral parameter, neurons distributed within multiple cortical areas participate in the representation and processing of that parameter. As an illustration, a study that employed multielectrode recordings from dorsal premotor cortex (PMd) and ventral premotor cortex (PMv) (68) showed that both areas represented the kinematics of arm reaching and characteristics of hand grasping, the conclusion that favored the distributed-coding principle and rejected a dual-channel hypothesis that attributed the representation of reaching to PMd and the representation of grasping to PMv (673).

Next, the neuronal multitasking principle proposes that an individual neuron can simultaneously represent multiple behavioral parameters, for example, arm kinematics and gripping force exerted by the hand (114). Consistent with this statement, a recent review by Fusi et al. (280) confirmed that neuronal responses often represent combinations of behavioral parameters. Fusi et al. (280) suggested that such high-dimensional representation is computationally advantageous compared with a population of highly specialized neurons because it allows the nervous system to generate a huge number of potential responses from the inputs to the

### Table 1. Principles of neural ensemble physiology

<table>
<thead>
<tr>
<th>Principle</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distributed coding</td>
<td>Representation of any behavioral parameter is distributed across many brain areas</td>
</tr>
<tr>
<td>Single neuron insufficiency</td>
<td>Single neurons are limited in encoding a given parameter</td>
</tr>
<tr>
<td>Multi-tasking</td>
<td>A single neuron is informative of several behavioral parameters</td>
</tr>
<tr>
<td>Mass effect principle</td>
<td>A certain number of neurons in a population is needed for their information capacity to stabilize at a sufficiently high value</td>
</tr>
<tr>
<td>Degeneracy principle</td>
<td>The same behavior can be produced by different neuronal assemblies</td>
</tr>
<tr>
<td>Plasticity</td>
<td>Neural ensemble function is critically dependent on the ability to plastically adapt to new behavioral tasks</td>
</tr>
<tr>
<td>Conservation of firing</td>
<td>Overall firing rates of an ensemble stays constant</td>
</tr>
<tr>
<td>Context principle</td>
<td>Sensory responses of neural ensemble changes according to the context of the stimulus</td>
</tr>
</tbody>
</table>

[From Nicolelis and Lebedev (583).]
neuronal population. Indeed, our neural degeneracy principle supports the hypothesis of Fusi et al. by proposing that a given neuronal ensemble can generate many behavioral outputs and, conversely, that a given behavioral output can be encoded by a variety of neural ensembles. According to our view, such a principle, which is similar to degeneracy observed in the genetic code (221, 853), improves the robustness and flexibility of neural encoding (480).

An additional clarification of the relationship between neuronal ensemble physiological patterns and associated behaviors is provided by the conservation of firing principle. This principle explains that, due to the metabolic constraint created by a fixed brain energy budget, the overall discharge rate of a neuronal population should always remain constant. Thus, if some cells increase their firing rate to encode a motor parameter, others should reduce their activity proportionally, to allow the energy consumption to remain around a set limit. This principle is closely related to the free energy principle which states that the brain minimizes its free energy when exchanging information with the environment (273–275, 806).

Within the constraints of the conservation of firing principle, the context principle states that information representations by neural ensembles depend on behavioral context. Neuronal response to an event would be different depending on the set of circumstances that surround the event, for instance, whether the animal is awake or anesthetized when the same sensory stimulus is delivered. Differences in neuronal firing patterns produced under these two different conditions, according to our view, reflect the brain’s ability to contextualize information (527). Finally, the plasticity principle states that neuronal populations modify their properties when an organism adapts to novel conditions or learns new behavioral tasks.

Our findings from BMI studies are consistent with these neural ensemble principles. From more than a decade of BMI research, we have learned that the best and most optimal way to extract a motor parameter from brain signals is to rely on concurrent recordings of the activity of large populations of neurons, distributed across multiple cortical areas. Clearly, this sample should include a large population of neurons located in M1, the cortical area from which most reliable kinematic and dynamic information can be obtained. Yet, surprisingly, many nonprimary frontal and parietal areas can contribute meaningfully to a BMI that continuously controls movements of an external actuator (3, 114, 852). On the basis of these findings, it is fair to say that BMI research has contributed decisively to consolidate a new view of cortical processing, one that departs from the classical dogma, proposed by the neuron doctrine, where a single neuron is considered the true functional unit of the brain, to one in which distributed neuronal populations assume that key physiological role. Put in other words, BMI research truly vindicated the Hebbian view of the brain.

Among multiple contributions to a better understanding of cortical ensemble physiology (583), BMI research has led to several demonstrations of new types of cortical plasticity. Indeed, since the early BMI studies, it became clear that it was only through cortical plastic adaptations that subjects could learn to control a BMI and improve their overall motor performance over time (114, 283, 377, 794). Eventually, this cortical plasticity led to the assimilation or incorporation of external actuators, such as robotic arms and legs, as if they were true extensions of the subject’s body representation that is known to exist in the brain (466, 583). BMI-associated cortical plasticity is manifested by changes in directional tuning patterns of individual neurons (114, 283, 463), alterations in temporal patterns of neuronal discharges (884), and a transient increase in the correlation between neurons within and between multiple frontal and parietal cortical areas (114, 377, 583). All these physiological adaptations mean that neuronal space in the cortex becomes devoted to representing a variety of properties associated with the artificial actuators employed by a BMI.

V. MULTICHANNEL RECORDING TECHNOLOGY

A. Microwire Recording Cubes

Developing methodology for reliable recordings from large populations of brain neurons is essential for further advances in our understanding of neuronal ensemble physiology and for the development of more practical BMI applications (466, 581, 583, 711). So far, the major achievements in this field have been associated with multielectrode implant methods.

We start by discussing the technology employed by our laboratory for the past 25 years because it has produced the highest neuronal yield and postimplant longevity reported in the literature so far (711). Thus, in our hands, a typical multielectrode implant was originally defined, in the mid-1990s, by a two-dimensional grid composed of small-diameter (12–50 μm) Teflon or isonel-insulated metal microwires (452, 581, 711). This technology evolved over nearly three decades of research, through experiments in rats, mice, monkeys, and human subjects. Optimal parameters were then obtained empirically over years of experimentation. Our latest configuration of this technology is defined by so-called volumetric, movable implants, named recording cubes, introduced by our laboratory over the past 3 years (FIGURE 2) (711). A recording cube is built by first constructing a grid of polyimide guiding tubes. The grid has a 10 × 10 arrangement with 1 mm spacing between the adjacent tubes. This spacing is optimal for monkey record-
ings; denser grids can be harmful to cortical tissue. The guiding tubes are contained in a three-dimensional printed plastic case that also holds an array of miniature screws utilized to move the microelectrodes. Each guiding tube accommodates 3–10 microelectrodes. These microelectrodes can have different lengths, which allow their placement at different depths in both cortical and subcortical areas. For that reason, these cubes allow us to obtain volumetric recordings from whatever cortical/subcortical structure is sampled.

During penetration into the brain, the microwires contained inside each tube move as a single set. This set includes the longest, leading microelectrode that penetrates the brain first. The leading microelectrode has a conical tip, whereas the other microwires in the set have either conical or cut angle tips. This configuration of the tip shape allows each microwire set to penetrate through the monkey pia. Each recording cube is compact (surface area of 0.22 mm² per recording channel) and light (11.6 g). Since a 10 × 10 tube grid can contain up to 10 microwires per tube, each recording cube provides a total of 1,000 potential recording sensors. We typically record 1–2 individual neurons per microwire, which provides a potential sample of 1,000–2,000 neurons per recording cube implanted. Since we usually implant 4–8 such recording cubes in a single monkey, the potential neuronal sample that could be recorded in each monkey ranges between 4,000 and 16,000 neurons.

Although in the past we experimented with a variety of microelectrode materials, for example, tungsten for the shafts and gold-plated tips (478), we currently use only polyimide insulated stainless steel microelectrodes, 30–50 µm in diameter for recordings in monkeys, because they were found to have the best longevity and quality of recordings. For recordings in rats, tungsten microwires are useful because they allow recordings for up to 6 mo and do not induce any significant neuronal death or tissue inflammation, although recording quality deteriorates with time due to glial encapsulation of the microelectrodes (271).

The quality of the implantation surgery is a key factor in defining the recording array performance and its long-term reliability (607). In our laboratory, primate surgery is conducted in strict sterile conditions under general anesthesia. The animal is placed in a stereotaxic apparatus, and craniotomies are made over the areas of interest. Dura matter is removed inside the craniotomy, and the guiding tubes are placed in light contact with the pia mater. The recording cube is then fixed to the skull with dental cement; stainless steel and ceramic bone screws serve as anchors. The implant is then encased in a three-dimensional printed protective cap, which can also house the components of our wireless recording system (FIGURE 2).

Microelectrode penetrations are performed 1–2 wk after the surgery under ketamine anesthesia. Generally, it takes 1–2 wk to insert 500–1,500 microelectrodes in each animal. The penetrations are performed slowly and over multiple days to minimize tissue damage. Importantly, during each penetration, only a small number of microelectrodes are moved to avoid dimpling of the cortex and a bed of nails effect, where an electrode array cannot penetrate because the pressure is evenly distributed among many electrodes. Depending on the recording cube design, rotation of one miniature screw brings in motion the microelectrodes located in 1–4 guiding tubes. Electrophysiological recordings are conducted simultaneously with each penetration to confirm that the microelectrodes gradually move through the cortical layers as the miniature screws are rotated. We usually penetrate the cortex with the microelectrodes spaced no closer than 2 mm, and perform penetrations with other microelectrode subsets on a different day. Once the microelectrodes are placed in their designated locations, they are never moved afterwards. Several months later, the guiding tubes are encapsulated by connective tissue, and eventually the spaces of the craniotomies may get ossified.

We usually place recording cubes over both hemispheres, including areas such as the primary motor (M1), primary somatosensory (S1), dorsal premotor (PMd), supplementary motor (SMA), and posterior parietal (PP) cortical areas. With this approach to cortical recordings, we have routinely recorded from 300–1,700 neurons in a single animal per day; good recording quality and high neuronal yield typically have continued for several years (FIGURE 16). Indeed, at the time of this writing, two of our monkeys completed more than 7 yr of cortical recordings after the original implantation surgery. The same microelectrodes can be used for both electrophysiological recordings and electrical microstimulation for many months (597–599). Viable recordings with implanted microwire arrays for 7 yr were also reported by the Rizzolatti laboratory (453).

We recently designed multielectrode implants suitable for chronic implantation into subcortical structures in rhesus monkeys, such as the neostriatum, thalamus, and the hippocampus. In this design, subsets of 10–30 microwires form bundles, staggered at 1–1.5 mm, which are inserted through individual guiding tubes. Each guiding tube is inserted into the brain at a depth of 5–15 mm (depending on the target structure). The microelectrode bundle is passed through the length of the guiding tube, and then travels an additional 5–10 mm to cover the area of interest.

B. Utah Array

While multielectrode implants employing individually movable microwire bundles can be placed at a wide range of cortical and subcortical depths, several laboratories have used an array constructed of 100 rigid microelectrodes in a
fixed arrangement, known as the Utah array or Utah probe (111). The array is micromachined from silicon. Each silicon needle is ~1.5 mm long. The needles’ shafts are coated with polyimide, whereas their sharpened tips are coated with platinum. The spacing between neighboring needles is 0.4 mm. Insulated gold wires make electrical contacts to the back sides of the needles.

Originally, attempts at slowly lowering the Utah array into cortical tissue failed because such a dense array produced cortical dimpling, due to the bed of nails effect, resulting in an incomplete penetration of some of the individual needles (111). To overcome this problem, the inventors of this device adopted an impact insertion procedure whereby the array is pushed into the cortex at a high speed, though a pneumatic gun (680, 681). An examination of the long-term recording performance of the Utah probe in the cat cortex showed that, 6 mo after the implantation, 40% of the needles could not record neuronal activity, most likely because of fibrous encapsulation (679). Indeed, extensive fibrous tissue was detected on the explanted probes.

Currently, the Utah array is the only microelectrode implant approved by the United States Food and Drug Administration (FDA) for human use. The array has been employed in several clinical studies that involved examination of epileptic patients (594, 808, 850) and BMI operation (3, 83, 152, 360, 361). Yet, instead of sampling single-unit activity, most of these studies employed recordings of multiunit activity defined as electrical signals that crossed a certain voltage threshold. The degree of contamination of this signal by electrical and mechanical artifacts was not reported, and the possibility of electrical cross-talk between the channels was not analyzed. In defense of the poor quality of recordings (single-unit recordings are considered a gold standard in the field), an argument was put forward that simple threshold-crossing is sufficient for BMI control (134). While this practical consideration could be valid for some implementations, the threshold-crossing method is prone to confusing noise and artifacts with real neuronal activity. That can easily lead to the undesirable outcome that artifacts contaminate the signal used for decoding and controlling artificial actuators. These are very real possibilities that could occur in clinical studies of BMIs, particularly

FIGURE 16. Number of neurons recorded by microwire implants over time. The graph shows the average number of units sampled per 32-channel connector. Data from eight monkeys are presented. Sample waveforms are shown for one of the monkeys (monkey M), for different dates after the implantation surgery. [From Schwarz et al. (711).]
in those in which human subjects, who suffer from quadriplegia, can still move their heads. Artifacts generated by head muscle contractions or overall head movement could be accepted by a threshold-crossing method as representing valid neuronal signals. Supporting this possible scenario, a curious example of controlling a BMI by EMG artifacts sampled by an EEG cap was recently reported (160).

Several biocompatibility issues have been reported for Utah arrays. The available data on human cortical tissue responses to Utah arrays indicate that these implants cause tissue reactions, such as microhemorrhaging, microglia activation, and long-term inflammation with the level of severity depending on the tissue damage during the implantation (251). Some of these unwanted effects could be caused by micro-movements between the brain and the needles. In addition to these biocompatibility issues, the Utah probe is suitable for recordings only from flat cortical surfaces, not from the sulci. For the flat cortical surfaces, recordings cannot be obtained from sites located deeper than 1.5 mm. Because of these shortcomings, the Utah probe cannot be considered as the final solution for human implants despite its current use in clinical trials. There is a growing consensus in the literature that research should continue into developing better recording technologies suitable for humans (251, 271, 391, 624, 646).

C. Improving Multielectrode Implants

While microwire arrays currently represent the most practical solution for large-scale, multi-area recordings from both cortical and subcortical structures, there is ongoing research into new technologies. One direction of this research is aimed at minimizing the micro-movements between the implant and the brain that could result in tissue damage and inflammation. One solution is to have the implant float with the brain instead of being anchored to the skull. In the design, pioneered by Gualtierotti and Bailey in 1968 (325), a lightweight implant is tethered by a flexible cable. Several implementations of this idea have been developed (563, 574, 758).

Another solution to decrease the motion between brain tissue and microelectrodes is offered by polymer-based microelectrodes that are more flexible than microwires and exert less strain on the brain tissue (4, 344, 451, 788). Such flexible arrays require specialized insertion techniques that temporarily stiffen the microelectrode shafts with biodegradable materials (344, 441) or polymers (754, 789), which are then dissolved in the tissue. Promising results were recently obtained using a sinusoidal probe anchored to a three-dimensional spheroid tip to minimize the movements between the probe and the brain (754). The other type of flexible electrodes contains magnetic materials (214, 383). These electrodes are inserted using an external magnetic field.

Several microelectrode designs have explored the possibility of placing recording points along an electrode shaft to increase neuronal yield and achieve recordings from a volume of the nervous tissue. This design is exemplified by the NeuroNexus array, also known as the Michigan Probe (23, 567, 822, 844). This is an array composed of silicon-based planar electrodes with multiple recording sites. The number of electrodes per shank and the number of shanks can be configured. The array can be used for recordings from local populations of neurons and for recordings from different cortical and hippocampal layers simultaneously (174, 548). Recordings with these arrays remain good during the early recording days, after which the recording quality usually deteriorates (409, 695).

Tetrodes represent a popular solution to improve the quality of single-unit isolation. A tetrode is composed of four twisted microwires with blunt tips (389, 664, 857). The microwires have differing lengths, which allows sampling neuronal potentials in a small three-dimensional space surrounding the microwire tips. Each neuronal waveform sampled by a tetrode yields a different amplitude for each microwire, due to the differences in distance from the neuron to each of the four microelectrode tips. This simple geometrical arrangement allows better sorting of single units based on the extra spatial sampling dimensions added to the process. Recently, tetrode designs have been applied to multi-area recordings in primates, where a microdrive advanced up to six guiding tubes containing tetrodes to several brain areas in awake rhesus monkeys, including primary motor cortex, prefrontal cortex, neostriatum, and hippocampus (693).

Ongoing research on new materials could further improve both microelectrode recording properties and biocompatibility. Keefer et al. (416) reported that coating of tungsten and stainless steel electrodes with carbon nanotubes improved the recordings by decreasing the microelectrode impedance and enhancing charge transfer. Suyatin et al. (779) developed an electrode based on gallium phosphide nanowires with the sensor composed of a deposited metal film. Overall, nanomaterials are a promising research area because they can provide better biocompatibility (66, 433, 446), spatial resolution (215, 799), and electrical properties (32, 118). Using pure carbon nanotube probes is another promising step in this development (880).

D. Neurotrophic Electrode

Neurotrophic electrodes were developed by Philip Kennedy in the late 1980s to early 1990s (418, 420, 422) as an endeavor to produce a long-lasting solution for the recordings of brain activity in paralyzed and locked-in patients to aid the communication of these patients with the external world. The idea was to make these electrodes biocompatible using neurotrophic factors that evoke growth of neu-
rites into the recording tip. The electrode was made of a hollow glass cone that contained three or four golden, Teflon-insulated wires glued to the cone walls. The cone was inserted in the cortex at a 45° angle to the surface. Histological analysis of rat (418) and monkey (420) implants showed that neurites grew into the cone and became myelinated. The tissue inside the cone contained axons, axodendritic synapses, blood vessels, and oligodendrocytes, whereas no microglial cells were detected. Bipolar recordings were conducted from pairs of wires. Kennedy and his colleagues reported that the cone electrode remained functional during the entire implantation period, which was 15 mo in the monkey (420, 422), 16 mo in the rat (418), and longer than 4 yr in humans (47).

Several BMIs were implemented with this implant. A patient suffering from brain stem stroke received the implant in an unspecified cortical area and was able to control a computer cursor with the recorded signals (419). The recordings continued for more than 17 mo. The patient had residual facial movements and eye movements with nystagmus. Correlation of the recorded signals with mouth, face, and eye movements were noticed over the 4 mo following the implantation. The patient stopped making these movements afterwards and learned to control the X direction of a computer cursor. The cursor moved to the right when the recorded signal increased, but did not move when the signal decreased. After the cursor reached the right edge of the screen, it returned to the leftmost position and shifted downward (i.e., a carriage return). With this simple control, the patient improved in two tasks: 1) moving the cursor to a screen icon and staying on it for 2 s to produce synthetic speech using phrases such as “Hello, my name is JR,” “I feel uncomfortable,” “I feel too cold,” “I feel too warm,” “Please help me,” and “I am in pain”; and 2) using a screen keyboard to spell phrases that are printed on the screen or vocalized by a speech synthesizer. The patient reached a spelling rate of three letters per minute. Kennedy and his colleagues claimed that the patient could dissociate neural activity from the facial EMG activity when controlling the cursor. It was also implied that the recorded signal was not contaminated by the EMGs and other artifacts.

In the next study, a neurotrophic electrode was placed in a speech-related area of the left precentral gyrus of a locked-in patient (327). The patient was paralyzed, with the motor output limited to the ability to produce slow vertical eye movements. The recorded potentials were transmitted using a wireless link to a speech synthesizer. The neuronal rates were converted into formants that represented small sets of continuous sounds (1-s long vowels “uh”, “iy”, “a”, or “oo”) using a Kalman filter. The sound served as auditory feedback. Aided with this feedback, the patient achieved a success rate of up to 70% on a three-vowel task.

Notwithstanding the significance of these results, the exact nature of electrical potentials recorded by the neurotrophic electrode remains unclear. From the information provided in these human studies, it is difficult to tell whether the signal contained neuronal spikes or field potentials picked by the microwires. To date, no other group reported using these or similar recording methods.

### E. Neural Dust

Neural dust is a recording method that utilizes small (10–100 μm) sensors (“dust”) that detect extracellular neuronal potentials and communicate them via an ultrasonic link to an interrogator placed under the skull (720). Each sensor contains a set of electrodes for recording neuronal activity, metal-oxide-semiconductor (CMOS) circuitry that amplifies the signal, and a piezoelectric transducer that converts electrical potentials into ultrasound. The interrogator uses ultrasound to both power the dust particles and examine their state. The interrogator also communicates with the extracranial components of the system.

While the main advantage of neural dust is the absence of microwire shafts that could be traumatic to the nervous tissue, the methodology has not been developed yet for injecting these particles into the brain. Additionally, fundamental concerns remain regarding effects of implantation, signal quality, separation of multichannel signals, and recordings longevity. There have been no recordings yet of cortical potentials using these devices, only recordings of large compound nerve potentials and EMGs (721). Therefore, the viability of this new method for cortical and subcortical recordings still needs to be demonstrated.

### F. Endovascular Electrodes

Using brain blood vessels as entry points for brain recording probes is an attractive possibility because this method could allow placing the recording sensors close to neurons without breaking the blood-brain barrier, by inserting thin electrodes in capillaries (507).

Endovascular stent electrodes have been used in cardiology for several decades. For example, Mirowski et al. (546) developed in 1980 an endovascular defibrillator for monitoring cardiac electrical activity and delivering defibrillating discharges when ventricular fibrillation was detected. Since that time, several modifications of such an endovascular cardiac electrode have been proposed (699, 700).

The same electrode insertion methods have been used for endovascular recordings of neural activity (84, 350, 712, 797). For example, Boniface and Antoun (80) recorded EEG activity using a Teflon-coated endovascular guide wire that was inserted in the middle cerebral artery in human.
epileptic patients undergoing preoperative carotid artery assessment.

Recently Oxley et al. (613) conducted multichannel recordings with an array of stent electrodes, called stentrodes, which they implanted into the superficial cortical veins overlying the sheep motor cortex. Blood vessels as thin as 1.7 mm in diameter were implanted with the stentrodes, yielding brain signals that were comparable to epidural ECoG recordings. Recordings in freely moving animals continued for 190 days. The authors proposed that such their endovascular system could be used to detect seizures in epileptic patients, operate BMIs, and deliver electrical stimulation.

Nanoscale electronics is a promising method for recordings from brain capillaries (507). Masayuki Nakao, Rodolfo Llinas, and their colleagues developed nanotechnology probes composed of insulated Wollaston platinum wires, 0.6 μm in diameter (507, 842). The feasibility of these recordings was demonstrated in the frog spinal cord. The sensors were introduced to the bloodstream through a polyimide tube, 90–300 μm in diameter. The wires “sailed” within a blood vessel until they straightened and could be used for recordings.

G. Optical Recordings

Optical imaging is based on voltage-sensitive (322, 323, 621, 793) and calcium-sensitive (320, 748, 771) fluorescent dyes, and genetically encoded calcium indicators (516, 523, 892) whose signals are monitored using video recordings. A particularly powerful method, two-photon excitation laser scanning microscopy, allows minimally invasive, three-dimensional sampling with submicrometer resolution (353, 591, 782). Although these techniques require filling neurons with a dye, hardly a practical procedure for human clinical applications, several optical imaging-based BMIs have been already demonstrated.

Clancy et al. (145) conducted experiments in mice with genetically encoded calcium indicator gCaMP6f in layers 2 and 3 in M1 or S1. Recordings were conducted using two photon imaging. Head-fixed mice were trained to control the pitch of a sound by modulating activity of neurons with a dye, hardly a practical procedure for human clinical applications, several optical imaging-based BMIs have been already demonstrated.

Ziv et al. (896) performed calcium imaging in freely behaving mice whose hippocampal tissue was virally infected and coexpressed GCaMP3 and CaMKII in the same neurons. A miniature (1.9 g) integrated fluorescence microscope (304) was employed for two photon imaging. While the mice explored their environment, place fields of thousands of hippocampal neurons were tracked over several weeks. Next, Bayesian decoding was employed to reconstruct the animal location within the environment from the optical recordings. The decoder was trained on the recording data collected on one day, and tested on the data the same day or different days. The decoding accuracy was the highest for the same day, but declined modestly for the different days. The authors explained these results by the dynamical changes in the hippocampal place fields.

H. Electrocorticographic Grids

ECoG grids, containing several tens to several hundred electrodes, allow minimally invasively recordings of multichannel field potentials from large cortical territories (172, 357, 483, 544).

A recent trend in this approach was to miniaturize the electrodes and increase their density (79, 542, 825, 838, 895), leading to improvements in spatial resolution of the recordings. The efficiency of recordings with high-density ECoG can be improved using electronics embedded in the grids. For example, Viventi et al. (825) developed ECoG grids with embedded flexible nano-membrane transistors that performed amplification and multiplexing. This technology reduced the number of connecting wires while increasing the number of recording channels to several thousands.

Fu et al. (278) developed flexible mesh electronics with micrometer components and bending properties matching those of neural tissue. The mesh consisted of 16 electrodes and could be injected in the mouse brain using a syringe. The implant yielded stable recordings of LFPs and single-unit activity for 8 mo. The same probe was used for long-term electrical stimulation.

I. Amplification, Processing, and Transmission of Neuronal Activity

A modern neuronal recording system typically includes several signal processing components. The preamplifier is usually placed near the subject’s head and is often called the headstage. It performs an initial amplification (typically with the gain from 1 to 20) and decreases the output impedance. This preamplification is needed to reduce the noise added at the next signal transmission stage. Depending on the system, the headstage output remains analog (601) or is digitized (135, 340, 711). The headstage output may also perform signal multiplexing for the reduction of the number of cables in a tethered system. The headstage is connected, via a tethered or wireless link, to an external processing unit that performs further amplification and/or filtering. Next, the neural signals are digitized if they were not digitized by the headstage. The digitized signals are then sent to a computer for further processing, including neural decoding for running BMI tasks.

While tethered systems were used in early BMIs (114, 124, 725, 794, 852) and are employed in human clinical trials (3,
83, 360), the increasing numbers of simultaneously sampled neural channels, the need to study BMIs while animals freely behave and the eventual goal of making BMIs fully implantable have prompted the development of wireless technologies. Figure 2 shows our recently developed wireless BMI, which has a 512–1024 channel capacity (711). This system includes four modules: 1) digitizing headstages housed in a monkey headcap, 2) wireless transceivers also housed in the headcap, 3) a wireless-to-wired bridge for bidirectional signal transfer, and 4) client software. Each headstage samples 32 channels, and each transceiver module connects to 4 headstages for a total of 128 channels handled by one transceiver. Up to four transceivers have been demonstrated to work simultaneously in rhesus monkeys. Recently, we found a solution to add 4 more transceivers for a total of 1024 channels. The bidirectional wireless link serves to transmit neural information to the external units and to set the headstage/transceiver parameters by the operator. Spike sorting operations are performed by the transceiver, which reduces the amount of neural information transmitted wirelessly. The headstage and transceiver are powered by a lithium-ion cell, which is housed in the headcap and operates for 30 h continuously. This wireless, multichannel recording system is connected to a BMI suite. The system has been already used in a variety of BMI tasks, ranging from BMI control of a cart by a monkey housed in its home cage (711), to a monkey performing BMI whole-body navigation tasks while seated in a motorized wheelchair (656). Several multichannel, wireless recording systems have been developed by other groups, as well (81, 119, 135, 340, 429, 550, 557, 602, 783).

VI. DECODING OF BRAIN SIGNALS

A. Principles of Neural Decoding

Modern real-time BMI computational algorithms, or decoders, are employed for transforming neuronal activity into signals suitable for direct communication of the subject’s brain with artificial actuators. Such decoders can employ a large variety of statistical and machine-learning methods. BMI decoding algorithms belong to the class of multiple-input and multiple-output (MIMO) models (432), where multiple inputs are provided by the neural recording channels and multiple outputs correspond to the behavioral variables controlled by the BMI and/or signals for communication with the external world.

A decoder applies a transform algorithm to neuronal inputs to calculate the output variables. In many cases, the transform algorithm has many independent parameters that need to be mapped to a much smaller list of output variables. Setting the values of these parameters is called decoder training. There are different methods to train a decoder. The most traditional one requires sampling an initial segment of input data from which correlations between neuronal signals and behavioral variables of interest are determined. The decoder performance is evaluated using the comparison of the actual behavioral parameters and the values derived by the decoder from the neuronal signals. After the decoder reaches high performance for the training segment, the BMI mode of operation can begin (851). At this point, the parameters needed to control an external device are derived solely by real-time decoding of the incoming neural activity.

The original BMI experiment performed by Carmena et al. (114) (Figure 5) can be used as an example of the traditional training paradigm to create a BMI decoder. The experiment started with monkeys using one hand to operate a joystick linked to a robotic arm. The joystick movements were translated into the reaching movements performed by the robot. Additionally, monkeys activated the robot gripper by squeezing the joystick handle. This hand-control phase lasted for ~15 min. Neuronal recordings obtained during this period provided the training data for a linear decoding algorithm, called the Wiener filter (see below), which represented the robotic arm kinematics and the gripping force as weighted sums of the neuronal discharge rates. Once the Wiener filter was trained, the operation was switched to BMI control, where the joystick was either electronically disconnected from the robot or in some experiments physically removed from the setup. At this point, the monkeys used the Wiener filter’s outputs to directly control the robotic arm’s reaching and grasping movements. This type of BMI operation was called brain control.

The shift from the training phase to brain control is often accompanied by changes in the patterns of overt movements produced by monkeys (114, 124, 463). Often, animals tend to diminish or eliminate their arm movements when operating a BMI in brain-control mode. Furthermore, the animals’ cortical neuronal firing patterns change significantly during the transition from the training period to brain-control mode. These physiological changes are manifested as increases in correlation between cortical neurons, within and between cortical areas, (114, 583), and alterations in the directional tuning of individual neurons (283, 463). If these changes occur, the decoder performance may deteriorate because it was trained under different behavioral conditions. To mitigate this problem, several strategies have been designed to adapt to these new behavioral and physiological conditions (114, 180, 283, 492, 609, 794). For example, the initial training period can be eliminated and the adaptation could begin from arbitrary decoder settings and continue throughout the experiment (283). Additionally, the initial training can be conducted without any overt movements, using only passive observations by the subject (377, 800, 827) and/or their mental imagery of movements (360, 362).

As mentioned above, after the original introduction of linear models for neural decoding in BMIs (374, 852), many other decoding algorithms have been proposed over the past decade. Indeed, describing a new BMI real-time com-
B. Linear Decoders

Linear decoders compute the output variables as weighted sums of the recorded neuronal rates (114, 794, 851, 852). Humphrey, Schmidt, and Thompson first proposed this idea in 1970 (374). They successfully demonstrated that movement parameters can be reconstructed from the recordings of the firing rates of multiple neurons using multiple linear regression. Schmidt and his colleagues then ran this linear decoding in real time (702, 703). Georgopoulos was a notable proponent of the theory that such weighted summation constitutes a fundamental mechanism by which cortical neuronal populations represent motor variables (295, 297, 299). Georgopoulos’ theory expressed the contribution of each neuron to population encoding as a vector that pointed to the so-called preferred direction of that neuron (294, 710). The preferred direction was defined as the direction for which the neuronal firing rate was maximal. Next, the individual-neuron unit vectors, multiplied by the discharge rate of the corresponding neuron, were summed to form a population vector. The population vector turned out to be an excellent method to continuously extract the direction of arm movement from the activity of a population of M1 neurons.

Since its introduction, the population vector approach to neural decoding has been very influential even though its original description was not based on simultaneous recordings from many neurons. Instead, Georgopoulos and his colleagues (295, 299) employed traditional single-electrode neurophysiology to record from a single neuron or, rarely, a few neurons at a time. Artificial neuronal populations were assembled from sequentially recorded neurons on different days, and population vectors were calculated for those populations. Although Georgopoulos later employed a seven-electrode apparatus for his motor cortical recordings (36, 173, 570), he did not record neuronal samples large enough to enable real-time decoding of arm movements using the population vector.

In the classical Georgopoulos paradigm, the neuronal preferred directions were derived from a behavioral paradigm called a center-out task. In this task, monkeys were required to perform arm reaching movements from an initial, central location, to a set of peripheral locations arranged in two- (294) or three-dimensional (299) space. Once the preferred directions were determined, the motor task could remain the same, or a more sophisticated task could be introduced. For example, Georgopoulos et al. (297) employed a cognitive task where monkeys had to perform a 90-degree mental rotation from the location of a visual stimulus to the motor target instructed by that stimulus. The population vector proved to be informative of the representation of mental rotation by M1 neurons even though the neuronal preferred directions were derived from a simpler, center-out task.

This early work on population-vector decoding contained an estimation of how many neurons would be needed to decrease noise in the extraction of arm kinematics from M1 ensembles (295). The decoding noise was expressed as the variability of the calculated population vector direction for different realizations of single-trial ensemble firings. This parameter was then plotted against the population size. The curve showed an initial rapid decrease in variability as the population size increased, and then followed a much slower rate of decrease after the population reached the size of 150 neurons. The slow decrease continued until all 475 neurons recorded in that study were included in the population. These estimations are consistent with our analysis of neuronal dropping curves constructed for simultaneously recorded populations of cortical neurons (467, 583).

Notwithstanding the elegance and theoretical importance of the population vector approach, this decoding method is suboptimal because the weights given to different neurons are chosen intuitively and without any provision for minimizing decoding errors. Additionally, the approach where the algorithm parameters are derived from one task (center-out task) and then applied to a different motor task (e.g., mental rotation) may result in additional errors because neuronal tuning properties could be different in the new task. Such a change in neuronal tuning is consistent with the context principle discussed above: under different conditions, cortical neurons tend to exhibit different activity patterns.

Another decoding algorithm, the Wiener filter, has been successfully employed in various BMI studies to extract limb kinematics and other behavioral variables from neuronal population activity (114, 283, 284, 432, 468, 656, 852). The Wiener filter is an optimal linear decoder set to minimize mean-square error (347, 485, 854). In a typical implementation, the Wiener filter output for time $t$ is computed as a weighted sum of neuronal rates sampled at several time points, called taps or lags, preceding $t$. Ten taps with 100 ms spacing is a typical setting in monkey BMI experiments (114, 463). The filter weights are computed by applying matrix transformations to the training data that include the recordings of neuronal rates and behavioral variables of interest.

The total number of Wiener filter weights depends on the population size and the number of taps. For neuronal population of size $N$ and number of taps $T$, the number of weights is equal to $N$ multiplied by $T$. The number of weights, or free parameters, is referred to as the dimensionality of the decoder. An excessive number of weights may be harmful for decoding accuracy because of the problem...
generated the robot’s gripping force. The same approach was subsequently used to extract kinematics of lower limb motion during bipedal locomotion, decode arm EMGs from cortical ensemble activity, decode cortical representation of time intervals, and control whole-body navigation by monkeys seated in a motorized wheelchair. Overall, this is a powerful and easily tractable decoding method.

### C. Kalman Filter

The Kalman filter (403, 404) is another popular algorithm for BMI decoding (432, 491, 620, 724, 869) that was previously employed in numerous engineering applications (319). Like the Wiener filter, the Kalman filter takes multi-channel neuronal signals as inputs and returns predictions of behavioral variables as its output. The filter models neuronal inputs as observations and the behavioral variables as state variables. The state variables may include, for example, the position and velocity of the arm.

The Kalman filter updates states in discrete time steps (usually 50–100 ms). Each update consists of two calculations. The first calculation, called predict step, provides an estimate of the next state from the previous state based on a state transition model, also called a movement model in BMI applications. For a robotic arm movement, for example, the next state can be estimated based on the previous position and velocity. The second calculation, called update step, performs an adjustment of this estimated state using recent states to the state variables. This enables the filter models neuronal tuning to speed in addition to velocity (492, 554). The UKF is relevant for neural decoding because there are nonlinearities in the relationship between neuronal rates and limb kinematics, such as the one relating neuronal rates to speed and limb kinematics.

In our implementation, an nth order UKF included two novel features: 1) a nonlinear model of neural tuning which incorporated absolute values of velocity and radius, and 2) an addition of n-1 recent states to the state variables. This new decoder outperformed the classical Kalman filter and the Wiener filter when applied to the data from a center-out task and a target tracking task. Moreover, the UKF outperformed the other decoders when used for real-time BMI control of movements of a computer cursor. Following this study, we have used the UKF in a number of BMI studies, including a BMI for controlling one avatar arm (598) or two avatar arms simultaneously (377).

### D. Point-Process Models

Point-process models of neuronal spiking activity employ a likelihood function to describe the probability of a neuron to produce a spike. The likelihood function depends on such parameters as the neuronal spiking history, activity of the other neurons in the population, external stimuli, and behaviors (144, 809). An analog of the Kalman filter can be formulated using a point-process model of the observation state (97, 222, 490). Indeed, both the Kalman filter and point-process decoders utilize the concept of state to describe both the decoded variables and neural activity. State transitions are described by statistical models (450).

Although computationally demanding, point-process decoders in certain cases offer better temporal resolution compared with neural decoders that decrease the resolution by down sampling the neuronal activity into bins (with a typical bin width of 50–100 ms) (461, 728–730, 807, 816, 872).

### E. Artificial Neural Networks

Artificial neural networks (ANNs) were introduced as decoders in the early BMI studies of the late 1990s (124, 852). Since then many ANNs have been used in both invasive and.
noninvasive BMIs, including a multilayer perceptron (22, 161, 369, 428, 431, 691), adaptive logic network (444, 445), tree-based neural network (380), and learning-vector-quantization (402, 460, 632).

Krishna Shenoy and his colleagues (776) have introduced a decoding method based on a dynamical ANN, called recurrent neural network (RNN). In this algorithm, neuronal activity is considered as a function of its history in addition to being related to motor parameters. In Shenoy’s study, the RNN continuously decoded the kinematics of center-out arm reaching movements from monkey M1 activity. This decoding scheme outperformed the velocity Kalman filter in the same task.

F. Adaptive Decoders

Adaptive decoders make adjustments that improve BMI performance while the subject continuously operates a BMI. While the decoder adapts, changes occur in the brain itself owing to neuronal plasticity.

The first adaptive decoder introduced in BMI literature was a coadaptive algorithm implemented to improve real-time conversion of the activity of monkey M1 neurons into the three-dimensional center-out movements of a cursor (794). The cursor position was generated by a population-vector decoder. The coadaptive algorithm adjusted the population vector weights after each trial to bring the BMI-generated trajectories closer to the ideal trajectories connecting the cursor’s initial position to the target. Following that study, many adapting algorithms have been developed.

Li et al. (492) developed a Bayesian regression self-training algorithm that updated the settings of a UKF. That BMI utilized neuronal ensemble recordings from multiple cortical areas in rhesus monkeys to control two-dimensional cursor movements. The adaptive algorithm monitored the decoder output and periodically updated the UKF neuronal tuning model based on the detected changes. The updates were performed using Bayesian linear regression. The online performance of this algorithm was tested in 11 experimental sessions that spanned 29 days. The initial parameters of the decoder were trained on the first day of recordings, and the evolution of these parameters was performed by the adapting algorithm without any retraining sessions. The adaptive decoder secured stable BMI performance, whereas the performance deteriorated if the unchanged initial decoder was used. Dangi et al. (181) used a similar two-dimensional reaching task as a test bed for developing a general framework for selecting parameters to adapt and the adaptation timescale. They also developed tools that evaluated convergence properties of adaptive algorithms.

Several studies introduced supervised learning algorithms that used the information about target location to adaptively improve BMI performance. Kowalski et al. (448) employed a naive adaptive BMI for this purpose. The system jointly analyzed neuronal patterns and user intent of reaching to the target. Shanechi and Carmena (728) used a similar idea of analyzing the user intent. They developed an adaptive optimal feedback-controlled point process decoder that derived subjects’ intentions from the relative position of the cursor and target. This adaptive scheme improved the performance even when the decoder’s initial parameters were set to arbitrary numbers. Along similar lines, Suminski et al. (774) developed a kinetic decoder that continuously adapted joint torques based on the discrepancy between the target location and hand position. The updates were performed using gradient descent.

Justin Sanchez and his colleagues (197) employed reinforcement learning as a BMI adapting algorithm. In this approach, actions that maximized the reward were selected through trial and error. Trial outcome (i.e., presence or absence of reward) served as a scalar signal that was utilized for parameter updates at the end of each trial. Two studies from this group (520, 640) showed that actor-critic reinforcement learning could quickly recover decoding accuracy when neural inputs were lost or shuffled. Moreover, the same group showed that the reinforcement signal for such learning could be derived from the recordings in nucleus accumbens (520).

G. Discrete Classifiers

Discrete classifiers convert neuronal activity into discrete choices. These decoders are commonly used in noninvasive BMIs, where subjects generate a limited number of outputs, often just two (333, 569, 639). Discrete classifiers have been used in some intracranial BMIs, as well (343, 692), including systems that combined both discrete and continuous decoders (261, 816).

The mathematical algorithms for discrete classification include linear discriminant analysis (LDA) (259, 288, 629), support vector machine (56, 288, 780), ANNs (359), multilayer perceptron (43, 75), hidden Markov models (603, 655), k nearest neighbors classifier (177), and nonlinear Bayesian classifiers (191).

VII. MOTOR CONTROL WITH INTRACRANIAL BMIs

A. Theories of Motor Control as a Foundation for Motor BMIs

Motor BMIs extract motor commands from a sample of neuronal activity and send this control information to external devices that execute the movements imagined by the operator. At the basic science level, these systems are in-
tended to investigate the physiological properties of motor circuits, theories on neuronal encoding, and the impact of learning and plasticity on neuronal ensembles. From a clinical point of view, BMIs primarily aim to restore crucial motor behaviors, such as arm movements or locomotion, to patients suffering from devastating levels of body paralysis, as a result of brain trauma or degenerative neurological diseases.

The design of existing motor BMIs, in many ways, matches current theories of the motor system layout and operation. The first design issue is where in the brain to record neural activity that would be converted into motor commands to an external device. This issue is closely related to the assessment of functional roles of different brain areas and the types of neural processing they perform. Since the motor system in primates is defined by a highly interconnected network of cortical, subcortical, and spinal structures, in general, there are many brain areas that could provide inputs for motor BMIs. Classically, the motor system is described as being formed by a hierarchy, in which cortical motor areas are presumed to handle advanced or higher order functions, for example, dexterous hand movements. Meanwhile, lower-order subcortical areas are presumed to manage less complex, automated motor acts. In this motor hierarchy, the spinal cord has been traditionally believed to handle low-order functions, such as reflexes (733) and central pattern generators (CPGs) (328). Reflexes are automated, and often unconscious motor responses to sensory stimuli. In contrast, voluntary movements are prepared and executed under cortical control. They may be related to external stimuli, but may also originate in the mind rather than being caused by sensory inputs. While there are merits to the classification of motor activities into less advanced, automated responses and more advanced, voluntary actions (65), our work has repeatedly shown that there is a constant flow of information between cortical, subcortical, and spinal structures during the execution of motor behaviors (576, 578, 617, 894). In this distributed view of the motor system, there is no clear-cut separation between high-order and low-order processing. In support of our view, practically any motor task involves a mixture of voluntary and reflex activities (158).

Several theories of motor control have influenced the design and experimentation with BMI. For instance, the concept of body schema, a quite important concept for modern BMI research as we discuss below, was originally proposed by Head and Holmes one century ago (351). According to Head and Holmes’ original formulation, the brain creates an internal model of the body, the body schema, which governs motor activities and perceptions. The body schema is constantly updated by streams of sensory information. With the emergence of BMI research, the concept of body schema was not only investigated but acquired a complete new angle; one in which artificial prosthetic limbs, controlled directly by the patient’s own brain activity during the utilization of a BMI, are believed to be assimilated into the body schema as extensions of the subject’s biological body, through the process of plasticity triggered by BMI long-term usage (466).

Modern theories of motor control are rooted in the ideas of Head and Holmes and, as such, have become the subject of investigation by BMI research. The internal model theory (310, 415, 459, 866) describes the motor system as being defined by two components: the controlled object (e.g., a body part or the entire body) and the controller. The controller uses an internal model to program future motor states. When the object movement is executed, the controller compares the expected state with the actual sensory feedback from the controlled object. If a discrepancy between the expected and actual state is detected, the controller issues a correction command. It has been suggested that to be efficient, BMIs should perform similar forward planning based on an internal model (175, 309).

Another popular motor control theory, Feldman’s equilibrium point theory (249, 250) suggests a possible neural mechanism to implement the controller. In this theory, higher-order motor centers manage the position of an equilibrium point for the limb, and the limb is brought to the equilibrium point by a spinal servo mechanism. BMIs with a similar separation between the higher-order and low-order controls have been proposed. In this design, called shared-control BMI, high-order motor commands are extracted from cortical activity, whereas the low-order execution is delegated to a robotic controller, which handles the “equilibrium point” using Feldman’s terminology (427, 634).

Optimal feedback control is yet another popular motor control theory (263, 801, 802). This theory describes an optimal strategy for using multiple biomechanical degrees of freedom to achieve the goal of a motor action. The strategy is based on stochastic optimal feedback control that corrects deviations in the degrees of freedom that define task goals, while allowing variability in task-irrelevant dimensions. The theory explains such phenomena as motor variability, error corrections, and motor synergies. Several BMI decoders that implement optimal feedback control have been proposed (59, 729, 730). These BMIs estimate the performance error by comparing the current location of an actuator with the planned trajectory estimated from the neuronal signals. A correction is then issued in the appropriate dimensions.

B. BMI Control of Virtual and Robotic Limbs

Motor BMIs that enable upper limb functionality, for example, a BMI for arm reaching and grasping (114) [FIGURE 5], have received particular attention because of the obvious
key importance of arm movements in our daily life. The first BMI of this type operated in an open-loop mode, i.e., without any sensory feedback from the BMI-controlled actuator (852). In that experiment, while New World monkeys manipulated a joystick, their cortical activity was decoded and converted into the movements of a robotic arm using an Internet protocol. The monkeys did not see the robot, which was located in a different state many hundreds of miles from the animal. All subsequent BMI demonstrations utilized a robotic arm operated in a closed-loop mode (114, 463, 794, 817), where monkeys received visual feedback of the robotic arm movements and could correct their performance errors.

The main components of a motor BMI that controls a robotic arm are featured in the system developed in our laboratory in 2003 (114) (FIGURE 5). Here, rhesus monkeys learned to control reaching and grasping movements performed by a robotic arm by using only the combined electrical activity of cortical ensembles recorded with multielectrode arrays, built from flexible Teflon-coated microwires chronically implanted in multiple cortical areas, including M1, S1, PMd, supplementary motor area (SMA), and posterior parietal cortex (PPC). These cortical areas were chosen because they belong to the frontal-parietal cortical circuitry that controls goal-directed arm and hand movements (20, 399, 470).

Two monkeys initially executed a motor task, placing a computer cursor to the center of a moving circle that served as a target, manually (FIGURE 5B). To do this, monkeys grasped a joystick and shifted it in different directions; the joystick position was translated into the cursor position on the screen. Later stages of this behavioral task also required that the monkeys apply gripping force to the joystick handle, at the end of the reaching movement, so that they could imitate grasping the virtual target. Next, monkeys learned to control the reach and grasp movements of the robotic arm equipped with a gripper. Since the joystick was connected to the robot arm, when the monkey moved the joystick and applied hand gripping force to it, the robot arm and gripper reproduced these movements. The visual feedback of the robot movements was delivered to the computer screen, where the robot position was represented by a circular cursor and the gripping force was represented by the cursor diameter. The virtual targets that the robot had to reach and grab were represented by circles of varying diameters. To win a fruit juice reward, monkeys had to move the robot, place its gripper over the virtual target, and then produce the correct level of gripping force, to match the cursor diameter with the diameter of the target.

While monkeys practiced these reach and grasp task, the firing rates of ~100 cortical neurons, distributed across the cortical areas mentioned above, were fed into multiple Wiener filter algorithms so that multiple parameters could be generated continuously to control the reach and grasp movements of the robotic arm. In brain-control mode, the joystick was electrically disconnected from the robot and the outputs of the Wiener filters defined the robot movements. The monkeys continued to manipulate the joystick with their hands although it was disconnected from the system. After the monkeys perfected this brain control assisted by the joystick movements, the joystick was removed from the setup. At this stage, to receive its fruit juice reward, the monkey could no longer rely on the well-trained joystick task. Instead, they had to learn to control the robot with their own cortical activity without assisting themselves with arm movements. The performance errors were initially high, but then decreased as the monkeys practiced in the brain control without hand movements. Learning to control this BMI was accompanied by a transient increase in correlation between the simultaneously recorded cortical neurons (114, 583), within and between cortical areas, and by changes in neuronal tuning to the robot arm movements (463).

Several other groups developed BMIs for arm reaching, as well. Schwartz and his colleagues have explored the possibility of performing BMI control over reaching in three dimensions (3D). In one study, they trained monkeys to wear stereoscopic goggles that displayed 3D movements of a cursor (794). In the beginning of each trial, the cursor was positioned in the center of this virtual reality display. A spherical target then appeared at a random 3D location, and the monkeys acquired it with the cursor to receive a reward. During the manual control mode, monkeys waved their hands in the air to move the cursor. The hand movements were monitored by a video tracking system. During the brain control mode, the cursor was moved by cortical activity processed by a decoding algorithm. Initially, the researchers attempted to train a population vector decoder using the manual performance data, and then use that decoder for brain control. However, after realizing that this control was not sufficiently accurate, they sought an adaptive algorithm that would improve the performance. Their adaptive decoder, called coadaptive movement prediction algorithm, adjusted the decoder parameters so that the trajectories generated by the BMI were brought closer to the ideal linear trajectories connecting the initial central position and the target.

Building on these results, the Schwartz laboratory developed a BMI for monkey self-feeding (817). For this purpose, they used a robotic arm equipped with a gripper that picked a piece of food and brought it to the monkey’s mouth. These experimental settings resembled the previous study of Lebedev and Wise where a robotic manipulator brought food to monkeys (471). In Schwartz’s study, the robot was controlled by a linear decoder that transformed cortical neuronal population activity into the velocity of the robot’s end point. The gripper’s opening and closing was commanded...
by cortical activity, as well. The authors reported a curious type of learning during these BMI operations: monkeys learned to start opening the gripper before it reached the target. They could do this without risking dropping a piece of food because marshmallows that stuck to the gripper were used as rewards. Although the monkeys possibly focused less on controlling the feeder because of the sticky rewards, this observation illustrates that BMI control, like normal motor control, can undergo adaptation.

Following these demonstrations of unimanual BMI control, the Nicolelis laboratory took the next logical step in this research by demonstrating a BMI that controlled two artificial arms simultaneously (377). In that study, rhesus monkeys viewed two virtual arms on a computer screen, and commanded their reaching movements using a cortical BMI that utilized the extracellular activity of ~400 neurons sampled in multiple areas of both hemispheres, including M1, S1, PMd, and SMA. Cortical ensemble activity was converted into bimanual movements using a UKF that treated the kinematic parameters of both arms as parts of the same state model. The decoder was trained either using a joystick task where monkeys moved the virtual arms with two joysticks or through a passive observation task that required monkeys to watch the virtual arms move on the screen. Eventually, the monkeys were able to control the virtual arms by their cortical activity without moving their own arms. This learning was accompanied by widespread cortical plasticity that manifested itself by an increase in cortical responses to the observation of virtual arm movements and by changes in pairwise correlations between neurons.

John Donoghue’s group at Brown conducted several BMI studies in implanted humans. In these studies, paralyzed patients were implanted with the Utah array in the M1. The patients learned BMI control of a screen cursor (361) or a robotic arm (361). One of the patients learned to grasp a coffee bottle with a robotic hand and, somewhat slowly (more than 1 min per trial), bring it to her mouth. The slowness of operation was possibly related to a deteriorated quality of neuronal recordings. The study did not document the number of neurons performing the control. Instead, it reported that neuronal electrical signals were picked up by 96 recording channels. A simple threshold crossing procedure was used to detect multiunit activity. The decoding was performed using a Kalman filter that was initially trained to predict robot hand displacements as the patients observed the movements of the robotic arm and imagined themselves controlling those movements. The Kalman filter decoder was iteratively adjusted during the phase of BMI control. To ease the learning, the patient’s performance was corrected by computer commands that brought the robot arm closer to the optimal trajectory. This procedure, called “error attenuation” consisted of decreasing the robot movement commands orthogonal to the trajectory connecting the robot to the target. The contribution from the error attenuation routine was gradually decreased and eventually removed. The robot hand state was controlled using an LDA classifier that, similarly to the velocity decoder, was trained using observations of the robot movements combined with motor imagery. The drinking task was further assisted by a preprogrammed sequence of actions. First, the LDA classifier commanded an automated impedance-controlled grasping and lifting the bottle. Second, the same classifier stopped the movements of the robot arm and pronounced the robot wrist to point the bottle toward the patient. Third, the robot wrist was brought to its initial position and arm movements allowed, and fourth, the bottle lowered to the table and released. Such a mode of operation, where control functions are distributed between a BMI operator and the robotic controller, is referred to in the literature as shared control (223, 379, 427).

While the experiments of Donoghue and his colleagues have demonstrated that patients with upper-limb paralysis can employ their cortical activity, recorded from the arm and hand representations in MI, to control the reaching and grasping movements performed by a robotic arm, several key questions remain regarding the nature of this control. The videos from their experiments show that the subjects could move their heads. In some of the trials, they clearly tracked the robot displacement with head movements. Arm movements accompanying the grasp command to the robot are also noticeable in the videos. These observations suggest that cortical neuronal activity related to head movements, which, in these patients, likely expanded beyond the original head representation before the trauma or disease, could have been involved in controlling the robot arm, in addition to the newly created cortical representation of the robot.

Overall, the role of assistive overt behaviors in BMI control, i.e., movements of body parts and the eyes that could be used to generate neural inputs for a BMI, is often neglected or downplayed in the literature. Certainly, this topic will require more scrutiny in the future. Historically, neurophysiological experiments strived for maximal control of unwanted overt behaviors. For example, neurophysiologists have developed an instructed delay task where an animal is not allowed to produce any motor output while preparing a movement (860). However, even if an instructed-delay task is well-learned, and no overt movement occurs, motor preparation still causes activation of spinal circuits involved in low-level motor control (653). Therefore, even a very clean BMI experiment, where the subject does not move the limbs or eyes, may involve activation of both higher-order brain areas that drive the BMI, and low-order subcortical and spinal regions. This is not a problem for practical BMI implementations, but rather an issue that needs to be better understood. For practical BMIs, even the presence of overt behaviors can be useful because they could improve the subject’s performance. Indeed, BMIs that mix several brain-derived signals with the signals representing
overt behaviors, for example, eye movements or EMGs, are called hybrid BMIs (337, 387, 476, 626, 833).

In their clinical trials, Schwartz and his colleagues further improved the accuracy BMI control exerted by a paralyzed human. They recorded from close to 200 neurons in an individual with tetraplegia implanted with multielectrode arrays in the motor cortex (152). As predicted in the early 2000s by our lab (114, 852), the increase in the number of simultaneously recorded neurons led to an improvement in the BMI performance. The subject gained control of an anthropomorphic robotic arm that performed skillful and coordinated reaching and grasping movements, like reaching to a knob and then turning it clockwise or counterclockwise. Like in Donoghue’s experiments, many of these experiments utilized assisted BMI control, where the subjects’ errors were corrected by the controller to facilitate learning. The subjects were eventually able to operate without that assistance.

Altogether, these clinical studies demonstrated the feasibility of implementing cortically controlled BMIs to reproduce upper limb movements. They also exposed a number of issues that preclude immediate translation of these systems into the clinical arena (466). One issue is the requirement for practical neural prostheses to be fully implantable. Wired implants are suitable for animal experiments and short-term clinical trials, but not for devices aimed at serving as long-term clinical solutions. In a practical clinical system, implanted electrodes and preamplifiers should be fully contained under the scalp while wireless technology is used to transfer large-scale recorded neural signals. Furthermore, implant biocompatibility remains a problematic issue. The utilization of the Utah probe, in both monkeys and human subjects, has repeatedly shown that the quality of neural recordings tends to deteriorate with time due to electrode encapsulation and neuronal tissue loss, likely as a result of the tissue injury caused by the electrodes. Finally, there are many challenges for real-time decoding algorithms, which currently are limited to small sets of motor behaviors.

C. BMI for Walking

During the last two decades of explosive BMI development, research focused mostly on controlling neuroprosthetic devices that mimic upper limb functions. Yet, tens of millions of people worldwide suffer from paralysis of the lower limbs as a result of trauma to the spinal cord or neurodegenerative diseases that affect the peripheral nervous system. Additionally, there are millions of lower limb amputees and patients who suffer from neurological disorders that affect gait, such as Parkinson’s disease.

A cortically driven BMI for decoding of bipedal walking was first developed by our laboratory (261) (FIGURE 12). In these experiments, two macaque monkeys were trained to walk bipedally on a treadmill while holding a bar with their hands to assist balance. Next, the monkeys were implanted with multielectrode arrays placed in the regions of M1 and S1 representing the lower limbs. The neuronal ensemble recordings conducted with these implants showed that, while monkeys walked on the treadmill, cortical neuronal discharges were correlated with the stepping movements. Owing to these correlations, the Nicolelis lab researchers could extract multiple lower limb kinematic parameters from the cortical recordings. Multiple Wiener filters were used for that purpose, which extracted 3D position of the hip, knee, and ankle joints, as well as the EMGs of leg muscles. The decoding reconstructed movement patterns of both forward and backward walking.

In a second series of experiments, using a custom-designed internet connection, the Nicolelis group transmitted the output of their BMI to a humanoid robot built by Gordon Cheng and Mitsuo Kawato at The Advanced Telecommunications Research (ATR) Institute in Kyoto, Japan (133). The humanoid robot received continuous signals from the BMI through an optimized internet link that minimized the transmission delay. An image of the walking robot was projected to the screen mounted in front of the monkey. Initially, the robot was suspended over a treadmill. In later experiments, monkey cortical activity was employed to induce controlled bipedal walking of the same robot on the floor (414).

After this study, decoding of kinematics of monkey quadrupedal walking from cortical activity was demonstrated by other groups, again with good precision (267, 268, 711). Additionally, leg EMGs during standing and squatting were extracted from monkey M1 activity (889).

For the case of quadrupedal locomotion, Capogrosso et al. (112) recently reported a “brain-spine interface” that alleviated gait deficits in rhesus macaques with unilateral spinal cord injuries. They implanted rhesus monkeys with multielectrode arrays placed in the leg area of M1, contralateral to the subsequent SCI site. Electrical stimulation was applied epidurally to dorsal roots to produce extensions and flexions of the leg weakened by the SCI. Monkeys learned to volitionally control the paralyzed leg using the interface that converted cortical neuronal activity into the spinal stimulation patterns. The authors argued that, because they recorded in M1 representation of the affected leg, the extracted motor commands represented intentions to move that leg. A careful examination of their methodology, however, raises several pivotal concerns. First, they employed a decoder training procedure that relied on the presence of overt movements in the affected leg that exhibited “residual hip or knee oscillations.” Clearly, this could be related to the mechanical perturbations caused by the movements of the intact limbs, rather than voluntary attempts to execute
steps with the paralyzed leg. Second, the neuronal modulations in the M1 ipsilateral to the lesion could represent the movements of the intact limbs that could move normally. Such representation could occur as the result of cortical plasticity following the SCI and maintained by cortico-cortical connections (393). Thus factors different from the monkey’s true intention to move the paralyzed leg could underlie the cortical modulations that triggered electrical stimulation of the dorsal roots and evoked the artificial steps.

In addition to assisting disabled people to regain the ability to walk, BMIs can be employed as a rehabilitation method (205). This latter approach, which has employed mainly noninvasive BMIs, will be discussed below.

D. BMI for Whole Body Navigation

Currently, the wheelchair is the main assistive device that enables navigation to people suffering from paralysis. Our laboratory also pioneered an intracranial BMI for wheelchair control (656). In this study, two rhesus monkeys were trained to control a robotic wheelchair, while being seated on top of it, by the activity of their cortical neuronal ensembles. Monkeys were chronically implanted with microelectrode arrays in multiple areas of both hemispheres. Neuronal ensemble activity in these areas was recorded using our wireless recording system (711). Wiener filters were used as decoders. Each experimental session started with a decoder training session, where the robotic wheelchair was driven by the computer; monkeys remained passive observers of these movements. During this passive navigation, two Wiener filters were trained to extract wheelchair kinematics from cortical activity. Such decoding was possible because cortical neurons were tuned to the wheelchair movements. One Wiener filter extracted translational velocity of the wheelchair (movements forward and backwards), whereas the other extracted rotational velocity (leftward and rightward rotations).

Following the training session, the mode of operation was switched to brain control, where the monkeys’ cortical activity was now mapped into the wheelchair’s translation and rotation velocities. The behavioral task consisted of driving the wheelchair toward a food dispenser that delivered grapes as a reward. As the monkeys trained, their ability to navigate the wheelchair with cortical signals improved. Additionally, performance on the wheelchair navigation task resulted in the emergence of a representation of the distance to reward location, a tuning property that resembled hippocampal place cells, in the primary motor and somatosensory cortical areas. This representation was totally unrelated to the settings of the decoder.

While our BMI converted M1 and S1 activity directly into whole body navigation commands, without the need for any intermediary overt behaviors, an alternative approach to enable such navigation is to have monkeys steer a motorized wheelchair with a joystick. Recently, it has been shown that monkeys can perform such steering to navigate a complex maze (226). Moreover, one study has shown a transition from joystick control to BMI control of a wheelchair (494). In that study, neuronal ensemble activity was recorded using cortical arrays implanted in the arm representation of M1. Monkeys were initially trained to steer the wheelchair with a joystick. While they did so, a decoder was trained to classify the joystick steering commands based on M1 activity. Next, the mode of operation was switched to brain control, where the steering command was derived from cortical activity. Finally, the authors demonstrated that, like in our study, the decoder could be trained without the joystick movements. Curiously, activity patterns of some M1 neurons changed dramatically after the mode of operation was switched from joystick control to brain control. Using the joystick in the context of BMI control of whole body navigation somewhat resembles the previous implementations of BMIs for arm reaching (114, 377). However, an important difference is that subjects have to learn a spatial transformation from the arm to the joystick movements (757).

Overall, these studies have demonstrated that intracranial BMIs could drive a prosthetic device that enabled whole body mobility. Such a device could be used to restore mobility to severely paralyzed patients in the future.

E. BMIs That Utilize FES

FES of peripheral nerves is a promising approach to restore motor functions to paralyzed subjects. FES-based BMIs aim to use the subject’s own brain activity to control the delivery of electrical stimulation to his/her own muscles that would then move their limbs. Over the past decade and a half, some progress has been reported with such BMIs.

The initial evidence of the feasibility of BMIs that mimic muscle activity was provided by the demonstrations that EMGs of arm (558, 642, 694) and leg (261) muscles could be extracted from the activity of cortical neuronal populations. Additionally, studies in healthy human subjects showed that a multichannel FES could produce near-normal hand movement patterns (390, 713).

The first demonstration of a BMI with FES output was done by Pfurtscheller and colleagues who aided a tetraplegic patient with a FES device attached to his forearm (630, 633). The FES was controlled by bursts of cortical beta activity (18–25 Hz) recorded by EEG electrodes placed over the patient’s sensorimotor cortex while he tried to imagine moving his foot. After some practice, the subject learned to grasp objects using this device.
Several demonstrations of BMI-controlled FES have been accomplished in monkey studies. Eberhard Fetz’ group temporarily paralyzed monkeys’ hands with an anesthetic blockade (536) and then employed the firing rate of neurons located in the primary motor cortex (M1) to control an FES device that evoked wrist torques. Visual feedback of the torque was provided by a screen cursor. Monkeys successfully learned to control this BMI. Moreover, M1 neurons, which were initially poorly associated with hand movements in a manual task, later on developed task-related modulations during the BMI control.

Lee Miller and colleagues also demonstrated brain control with an FES device (225, 641). Their FES system was controlled by populations of ~100 motor cortical neurons recorded with chronically implanted microelectrode arrays in the monkey M1. Hand paralysis in these animals was induced by an anesthetic block of the median and ulnar nerves at the elbow level. After the voluntary motor control of the hand was extinguished, the researchers activated forearm muscles with FES driven by the M1 signals. Monkeys were able to perform object grasping with this neural prosthesis.

Bouton et al. (83) demonstrated a BMI with FES in a paralyzed human with an intracranial multielectrode implant. The study subject suffered from a C5/C6 complete, nonspastic quadriplegia, resulting from a diving accident. A Utah array was implanted in the hand area of M1, which allowed recordings from up to 50 single units simultaneously. The recordings continued for 350 days, and 33 units were isolated by the end of the study. During the training session, the subject attempted to produce six wrist and hand movements. These movements were impaired by the paralysis but could be evoked by FES. The FES was delivered using a 130-electrode array of surface electrodes embedded in a sleeve that was wrapped around the forearm. Neuronal population activity was converted into FES patterns using multiple simultaneous neural decoders based on a nonlinear kernel method with a non-smooth support vector machine. During the brain control mode of operation, the subject was required to generate hand movement that matched the cue shown on a computer screen. Following training, he managed to perform up to 70% of trials correctly.

Notwithstanding the success of these demonstrations, using FES to restore movements meets a number of difficulties, such as muscle fatigue (224, 305, 796) and difficulties in achieving good accuracy of evoked movements without sensory feedback of force and position (25, 385, 818).

F. Neuronal Plasticity in Motor BMIs

Subjects usually experience difficulties when they are first introduced to brain-control mode of BMI operation. Yet, over time, they improve their performance with continuous practice. Such improvements have similar mechanisms as learning of new motor skills (1, 69, 212, 213, 356, 439, 460, 547, 726). As such, many authors have proposed that neuronal plasticity is essential for BMIs to work properly in both animal and human subjects (114, 170, 193, 204, 324, 463, 466, 583, 612). As a matter of fact, some authors have gone as far as to implicate cortical plasticity triggered by BMI operations as the key mechanism through which subjects could assimilate prosthetic limbs or even other actuators, such as virtual limbs, as extensions of the subject’s body schema created by the brain (466, 586, 738).

In general, BMI control of an artificial actuator has much in common with the neurophysiological mechanisms involved in learning to use and become proficient in tool handling, operations known to evoke brain plasticity (67, 195, 378, 524, 525). This likeness can be easily verified by reviewing the experiments conducted by Atsushi Iriki’s laboratory. In their fundamental experiments on primate tool usage, Iriki et al. (378) trained macaque monkeys to reach toward distant objects, which could not be accomplished by using their arms alone, by utilizing an external tool: a rake. Before monkeys could use the rake, the researchers measured the receptive fields of multimodal neurons in the posterior parietal cortex. Prior to the use of the artificial tool, these parietal cortical neurons exhibited both tactile and visual receptive fields (RFs) related to the animal’s hand: while the tactile RF was located on the hand skin, the visual RF was circumscribed to the visual space that closely surrounds the hand, the so-called peri-personal space. After the monkeys practiced and became proficient in the task of retrieving grapes with the rake, Iriki et al. observed that the visual RFs of the parietal neurons expanded to include the entire length of the rake, in addition to the peri-personal space around the animal’s hand. The Iriki laboratory interpreted these results as a suggestion that these cortical adaptations represented modifications of the animal’s body schema that resulted in the incorporation of the rake as an extension of the animal’s arm, as seen from the brain’s own point of view (586).

Long-term operation of BMIs that control the movements of artificial actuators, robotic or even virtual arms, leads to similar brain remapping of the receptive fields of cortical neurons located in multiple motor and somatosensory areas, as described by Iriki in their experiments with tool usage. Several studies from our laboratory reported neuronal plasticity during learning to operate BMIs, starting with the study by Carmena et al. (114) that showed changes in neuronal tuning curves accompanied by changes in correlation between neurons as monkeys learned to operate a BMI that enacted reaching and grasping movements. Changes in neuronal tuning were further investigated by Lebedev et al. (463), and Zacksenhouse et al. (884) reported stronger cortical firing modulations during learning of BMI tasks, which decreased after monkeys learned. Transient increases in correlations between neurons, associated with learning a
bimanual BMI task, were confirmed by our laboratory (377). Overall, these studies showed that, after the mode of operation is switched to brain control, neuronal activity patterns markedly changed, both at the level of individual neurons and their populations. Changes in neuronal tuning were observed even when monkeys continued to perform arm movements during brain control. In this case, neuronal tuning to movements of their own arms weakened, and the neurons started to represent the BMI-controlled actuator instead (463). Moreover, neuronal tuning to the actuator remained even when monkeys stopped moving their own arms (114, 377, 463). At the population level, switching to brain control was associated with increased synchrony between the neurons and, consequently, with many neurons having very similar preferred directions (114, 377, 583, 598).

Evaluation of changes in neuronal tuning during BMI operations has several caveats. The main factor that should be considered in such an analysis is that, during brain control, neuronal tuning properties no longer depend on the brain circuitry alone, but also essentially depend on the decoder settings. Indeed, the decoder uses a transfer function or an algorithm to translate the activity pattern of each neuron into a contribution to actuator movements. In the case a Wiener filter is used for decoding, the contribution of a neuron to a given degree of freedom is defined by the weight assigned to that neuron. For example, if a neuron is set to have rightward tuning, the decoder translates the discharge rate of that neuron into increments of the $x$-coordinate of the actuator, while no contribution is made to the $y$-coordinate. This decoder-assigned tuning may be different from the true neurophysiological properties of the neuron. Say, the neuron has switched to representing the $y$-coordinate and now matches the user’s intention to change the actuator’s $y$-coordinate and/or responds to the visual of the $y$-coordinate. Despite this new representation, the neuron will still contribute to the $x$-coordinate only because of the original decoder settings. In another scenario, the neuron fires at random and does not represent any intention or feedback, but still has a directionally tuned contribution established by the decoder.

The interpretation of neuronal tuning during BMI control is further complicated by the ensemble properties of the neurons contributing to the decoding. Consider a Wiener filter (for simplicity with just one tap) applied to a population of randomly firing neurons. An assessment of directional properties of each neuron in the population would show cosine tuning with a preferred direction defined by the Weiner filter weights assigned to the $x$ and $y$ dimensions. In this example, neuronal tuning during BMI control does not necessarily match any neuronal representation of the user motor intention and/or sensory feedback from the actuator; the tuning only corresponds to the decoder settings. The next level of complexity to the analysis of tuning properties of a neuron during BMI control is brought by interferences from the other neurons. The BMI output is produced by many neurons, not just by the neuron whose tuning is being assessed. Therefore, when the firing of one neuron is compared with the BMI output, for example, cursor trajectory, this is effectively a comparison of activity of one neuron with a variable composed from the activity of many other neurons. Consequently, the tuning assessment critically depends on the relative contribution of different neurons. Two extreme cases can be considered: 1) the neuron in question has a very strong contribution to the decoder output whereas the contribution of the other neurons is relatively small, and 2) the neuron’s contribution is very small and the decoder output is dominated by the other neurons. In the first case, the neuron’s tuning will mostly represent the decoder settings for the reasons explained above. In the second case, the neuron’s tuning will reflect the relationship of that neuron’s firing to the actuator position generated by the other neurons, irrespective of the decoder weights assigned to that neuron. Furthermore, correlated firing between neurons may have strong effects on the BMI output and consequently on the tuning of individual neurons. For example, if activity of a weakly contributing neuron is correlated with the activity of strongly contributing neurons, the tuning properties of the former will be very similar to those of the latter. Overall, although characterizing neuronal patterns during BMI control using tuning curves is helpful to reveal some basic features (114, 317, 463, 598), interpretation of such tuning characteristics is not trivial.

The pitfalls of neuronal tuning analysis for BMIs can be illustrated by the study of Ganguly and Carmena (283) that attempted to characterize the formation of new “cortical maps” as the result of learning to control a BMI. In that study, monkeys performed a two-dimensional center-out task using a BMI based on the recordings from small (~15 neurons) M1 ensembles. The small-ensemble activity was translated into cursor position using a Wiener filter with 10 taps. The study claimed that if the decoder is trained on day one and fixed afterwards, M1 neurons would plastically adapt to improve BMI performance and form a “cortical map.” The authors argued that the same M1 ensemble could simultaneously hold several “cortical maps” corresponding to different decoder. The “cortical map” was defined as a set of directional tuning curves, one per neuron. A close examination of this analysis reveals that neuronal tuning was determined differently from how it was set by the decoding algorithm. The 10-tap Wiener filter (100 ms bin width) effectively assigned 10 tuning curves for each neuron, one per tap. Yet, the authors chose to compute one tuning curve per neuron, which was derived either from a relatively short time window (200 ms) or a long one (2 s). The tuning curves were normalized to change from $-1$ to $1$, which made it impossible to compare tuning strength in different neurons. Factors like the relationship of these tuning curves to the fixed decoder settings, relative contribu-
tion of different neurons to BMI output, and neuronal correlations were not considered. As training on the BMI task continued for 9–19 days, the tuning curves changed during the initial training days and later stabilized, which was interpreted as the formation of a “cortical map.” These changes were paralleled by a clear evolution of cursor movements, which started as highly convoluted trajectories resembling random walk, but changed to almost straight center-out trajectories during the late training days. This meant that the early and late tuning curves were generated from very different cursor movements, which by itself can explain the differences that appear as changes in neuronal tuning. For example, the 2-s time window most definitely contained movements in very different directions during the early days, and represented a more uniform sample during the late days, an analysis that is guaranteed to generate different looking tuning curves. Therefore, the seemingly paradoxical result that neuronal tuning curves changed for a fixed decoder that presumably would have kept them very stable, most likely reflected changes in cursor movement patterns rather than any meaningful characteristics of neuronal representation of the external actuator. Given these considerations, that study’s conclusion regarding the emergence of a “stable cortical map” appears questionable. A more plausible conclusion is that both BMI output and the underlying neuronal patterns changed during learning. More data would be needed to evaluate if there was any change in the cortical representation of the actuator movements resulting from learning to control the BMI. Specifically, answers to the following questions would be needed: 1) how the neuronal tuning curves are affected by the decoder settings; 2) what other factors affect the tuning besides the decoder settings; and 3) how the tuning characteristics could be compared for datasets with very different actuator trajectories.

In the above examples, the major difficulty in evaluating neuronal tuning during BMI control is related to a somewhat circular approach: the actuator position is first generated from neuronal activity using a mathematical algorithm, and then an attempt is made to determine the relationship between the neuronal patterns and actuator movements once again, and to extract the features in this relationship that are not explainable merely by the decoder settings. This difficulty can be avoided if neuronal tuning is assessed based on parameter that is not generated by the decoder and can be manipulated independently of the decoder settings. Such an analysis was conducted in our study of a BMI for bimanual movements (377). In that study, we evaluated neuronal tuning to target position, the parameter that unrelated to the decoder settings. Monkeys controlled 2D movements of two virtual hands using a BMI; a separate target was designated for each hand. Since the target positions were not included in the decoder variables, neuronal tuning to the targets could not be a consequence of the decoder settings. A k-nearest neighbor (k-NN) classifier UKF decoder was used to extract the screen locations of targets from cortical ensemble activity on each behavioral trial, and the percentage of correct classifications was used as a measure of representation strength. The locations of the targets for both virtual hands were clearly represented by the cortical neuronal ensemble. These representations persisted during brain-control trials and passive observation trials. The passive observation trajectories did not change day to day. Therefore, we used them to assess long-term changes in the neuronal responses to the virtual hands. This analysis was valid because passive observation trials did not involve BMI control. The analysis was conducted offline by applying a UKF decoder to the neuronal recordings and using decoding accuracy as a measure of tuning strength. We found a clear improvement in decoding accuracy across the training days.

Several studies evoked learning (and related plasticity) by altering BMI decoder settings and observing behavioral and neural adaptations to such manipulations. Thus Chase et al. (127) examined a BMI that generated 2D cursor position from monkey M1 activity using a linear decoder (127). Next, they applied a rotational transformation to the contribution to the BMI output from a subset of neurons. Although this manipulation initially resulted in curved cursor trajectories, their monkeys adapted to the new condition and straightened the trajectories. The analysis of neuronal responses showed that the entire neuronal population contributed to that adaptation, not only the neuronal subset with perturbed BMI outputs. Using a similar manipulation, Ganguly and Carmena (283) perturbed BMI output by randomly shuffling neuronal inputs to a fixed Wiener filter. Their monkeys successfully adapted to that perturbation.

Sadtdler et al. (684) devised a method that made adaptation to BMI control particularly difficult. They applied a factor analysis to extract correlated neuronal responses and represent them as an intrinsic manifold, a subspace in a multidimensional space of population firing rates. The authors found that monkeys successfully learned to control the BMI with the inputs taken from the manifold, but learned with great difficulty if the inputs came from the outside of the manifold. In other words, monkeys adapted to a new decoder if it did not require them to alter the original structure of neuronal correlations. Although this study seems to suggest an existence of strong synergies between the neurons in an ensemble, there is also an alternative explanation. The study utilized a threshold crossing method for detecting multiunit spikes, a method prone to inclusion of noise into the spike data. The noise most likely ended up outside the intrinsic manifold, so in the outside-of-manifold task monkeys were asked to control the BMI with noise, obviously a task impossible to learn.

In addition to brain plasticity induced by learning to operate motor BMIs, plasticity occurs after training with sen-
sory BMIs (see the sections below on sensory and bidirectional BMIs). Studies conducted in our laboratory (341, 798) enabled rats to perceive infrared light using a BMI that converted the signals from the head-mounted infrared sensors into ICMS of the rat primary somatosensory cortex (S1). Learning to use this BMI resulted in the emergence of a representation of infrared light in S1. Moreover, this new representation coexisted with S1 representation of the rat whiskers.

Notwithstanding the progress in studies on BMI-induced neuronal plasticity, this research is still at the initial stage. Much more work will be needed to further clarify our understanding of this phenomenon. Yet, if one would have to choose one consensual view in this field, it is the assumption that without the occurrence of some level of cortical plasticity, BMIs would not be able to operate as successfully as they do. In other words, after a decade and a half of intense work, BMIs certainly owe their prominence in systems neuroscience to the exuberant propensity of the adult mammalian cortex to adapt itself to new task contingencies, particularly when exposed to rich feedback signals.

VIII. NONINVASIVE BMIs

A. EEG-Based BMIs

EEG-based systems are the most popular noninvasive BMIs, which have been thoroughly studied in humans, in both healthy subjects and patients. While approaches to EEG decoding are somewhat different from those used for extracting motor commands from streams of neuronal spikes, the general principles of neuronal ensemble physiology (583, 586) still apply to these applications. For example, decoding accuracy improves when more EEG channels are added (141).

Although EEG signals are prone to be contaminated with many sources of noise, including facial EMG, electrooculogram (EOG), and all sorts of movement artifacts, and despite the fact that EEG recordings do not yield detailed motor information compared with intracranial single-unit recordings, EEG-based BMIs have been successfully implemented in both normal and disabled human subjects to enact motor commands and provide communication channels. In particular, these EEG-based BMIs have been extremely useful in allowing “locked in” patients, those suffering from a complete level of body paralysis, as a result of a neurodegenerative disorder such as amyotrophic lateral sclerosis (or Lou Gehrig’s disease), to regain the ability to communicate with the external world, using EEG based spelling devices (71–73, 129, 184, 358, 364, 455, 481, 573, 606, 635, 718, 740, 742).

The BMI classification into independent (endogenous) or dependent (exogenous) systems (see sect. III) is particularly distinct for EEG-based systems. In independent BMIs, subjects perform volitional mental tasks, for example, motor imagery, that evoke changes in their EEG rhythms (3, 5, 17, 27–29, 31, 51, 77, 408, 560, 611, 632, 654, 811, 813, 826). The EEG bands typically employed in such BMIs are slow cortical potentials, mu (8–12 Hz), beta (18–30 Hz) and gamma (30–70 Hz) waves (14, 74, 815, 863). The majority of EEG-based BMIs translate EEG activity into discrete choices (17, 110, 159, 355, 370, 667, 669, 722, 840), but continuous control is also possible with independent EEG-based BMIs (211, 864, 893).

Dependent EEG-based BMIs utilize computer screens or LED displays as sources of visual stimuli that evoke EEG responses. Users modulate these responses to produce a BMI output (10, 42, 70, 74, 101, 117, 521). Classification algorithms identify cortical responses to screen stimuli to which participants attend (overtly or covertly). For example, a decoder based on visual evoked potentials (VEPs) (162, 196, 216) exploits the fact that VEPs are stronger in the visual areas when subjects attend to the stimulus and/or look at it (147, 216, 555, 566). The P300 component of the response to a visual stimulus, also called P3 (206, 637), has been particularly popular in BMI designs (10, 15, 16, 34, 54, 209, 243, 246, 257, 285, 363, 406, 521, 538, 545, 590, 717, 735, 736, 787, 832) because of its high sensitivity to subjects’ reaction to the stimulus in “oddball” paradigms, where a person is required to detect a target within a train of irrelevant stimuli (12, 206–208, 778). Auditory-based (58, 332, 410, 440, 455, 643) and tactile-based (95, 514, 814) P300 interfaces have been implemented, as well. Another popular BMI design is the one that utilizes VEPs generated by rapidly occurring stimuli (up to 60 Hz) (8, 42, 117, 122, 186, 187, 386, 395, 473, 481, 502, 515, 868, 891). Such VEPs are called steady-state visual evoked potentials (SSVEPs). In a typical SSVEP-based BMI, multiple flickering objects are shown on the screen; the flicker frequency is unique for each object. Subjects look at a specific object to issue a BMI command.

Hybrid schemes for EEG-based BMIs have also been developed (561, 626). Such BMIs combine several decoding principles, for example, motor imagery combined with SSVEPs (13, 100) or steady-state somatosensory evoked potentials (7), functional near-infrared spectroscopy (fNIRS) with asynchronous sensorimotor rhythms, P300 with SSVEP (154), and eye position with EEG decoding (337, 387, 833). Such hybrid BMIs are more versatile and accurate compared with BMIs that use only one control mode.

Overall, EEG-based BMIs have successfully achieved many significant milestones. These include spelling devices (10, 71, 101, 117, 132, 438, 559, 608, 619, 672, 876, 882), speech generators (190, 326, 549), BMIs for control of a humanoid robot (55, 103, 120, 138, 301), telepresence systems (130, 223, 804), BMI-controlled wheelchairs (121,
Notwithstanding these successes, it has been noted that many publications on EEG-based BCIs do not contain information on how EEG artifacts were handled (244). This is a serious problem because artifacts not only contaminate EEG recordings, but they could serve as a source of control signal for a BCI. Curiously, users can easily utilize their facial EMGs, produced by clenching the teeth and recorded with regular EEG electrodes placed on the scalp, to control a robotic arm. In fact, this EMG control outperforms EEG control in the same settings (160).

While EMG artifacts can be partially filtered out from EEG recordings by removing high-frequency signals (46, 311), mechanical artifacts strongly affect the EEG low-frequency range (331). Low-frequency mechanical artifacts can jeopardize the performance of BCIs that are based on slow cortical potentials. For example, José Contreras-Vidal and his colleagues employed linear regression decoders to reconstruct three-dimensional hand kinematics (86) and leg kinematics during treadmill walking (652) from slow cortical potentials. No artifact removal was performed, which opens a possibility that mechanical artifacts could influence the reconstruction. Strong objections to these results were expressed by Castermans et al. (116) who showed that EEGs recorded during treadmill walking were contaminated by the harmonics of the stepping frequency. In addition to questioning the reconstruction of movements from a low-delta EEG band, this study found that mechanical artifacts covered a wide range of EEG frequencies, so artifact removal by frequency filtering appeared to be unreliable. Yet another study (33) claimed that the employment of linear regression methods to reconstruct movements from cortical slow potentials was statistically invalid, since similar reconstructions could be obtained from both real and random EEG data.

### B. Magnetoencephalography-Based BMIs

In addition to electrical fields emitted by the brain, noninvasively recorded brain magnetic fields have also provided signals for BMIs (104, 413, 458, 541). Magnetoencephalography (MEG) detects weak magnetic fields produced by the electrical currents generated mainly by cortical neurons (149, 150, 334). Brain magnetic fields are very weak, on the order of picoTesla, or 10 million times less than the Earth’s magnetic field. To detect these tiny signals, one needs to employ very sensitive magnetometers. Historically, the copper induction coil was the first magnetometer (150), later replaced by superconducting quantum interference devices (SQUIDs) (149). Recently, atomic magnetometry was introduced based on a spin-exchange-relaxation-free (SERF) magnetometer that is more sensitive than SQUID (871). MEG has better temporal and spatial resolution compared with EEG, but they can be used only in magnetically shielded facilities.

Salmelin and Hari (688) reported suppression of the mu-rhythm, recorded with MEG, by thumb movements. Georgopoulos and his colleagues (296) employed a 248-sensor MEG to reconstruct arm movements from MEG recordings using a linear decoder. The first real-time MEG-based BMI was developed by Lal et al. (458); the system applied a binary classifier to MEG mu-rhythm. Mellinger et al. (541) used a similar design of a MEG-based BMI to classify motor and beta rhythms. MEG-based BMIs have been implemented in tetraplegic (413) and stroke (104) patients.

### C. fNIRS-Based BMIs

fNIRS applies light in the near-infrared range (600-1,000 nm) through the skull to detect changes in oxyhemoglobin and deoxyhemoglobin concentrations in the brain blood flow (252, 388). fNIRS-based BMIs have been gaining popularity recently (366, 533, 568, 744). fNIRS measures cortical metabolic activity with a spatial resolution of ~1 cm and temporal resolution on the order of 100 ms. However, the delay between neural activity and blood oxygenation changes is several seconds.

The first fNIRS-based BMI was demonstrated by Coyle et al. in 2004 (166). That BMI extracted brain signals related to motor imagery when patients were asked to squeeze a ball. For this purpose, a single optode was placed over the motor cortex representation of the patient’s hand (EEG coordinate C3). Increases in oxyhemoglobin and decreases in deoxyhemoglobin were detected when subjects imagined the contralateral hand movements. The neurofeedback generated by the patient’s motor imagery was provided by a variable-diameter circle shown on the screen. The BMI had 75% accuracy in recognizing the patient’s imagined movements. Approximately 5 s were required to register the change in the hemodynamic response.

Sitaram et al. (746) employed multiple optodes (four illuminators and four detectors for each hemisphere) placed over the motor cortex. The subjects performed finger tapping with the left hand or right hand, or imagined these movements. Support vector machine (SVM) and Hidden Markov Model (HMM) algorithms allowed this BMI to achieve greater than 80% accuracy in decoding which finger, right or left, was moving.

A number of studies employed fNIRs to recognize prefrontal cortex activity related to performing mental arithmetic (52, 53, 375, 649, 650). Additionally, fNIRs of prefrontal cortex activity was employed to decode subjective preference (517), music imagery (241, 649, 650), and emotional states (786).
As mentioned above, fNIRS can be used in combination with other neural recording methods to create hybrid BMIs. For example, hybrid EEG-fNIRS BMIs have improved the speed of fNIR thanks to the EEG’s superior temporal resolution. In addition, this hybrid method has allowed the generation of a larger repertoire of commands because different types of brain-related signals are employed in the discrimination performed by this BMI. For example, Fazli et al. (248) recorded cortical sensorimotor rhythms simultaneously with EEG and fNIRs methods, which resulted in better classification of motor imagery. Khan et al. (424) positioned the fNIRS sensors over the prefrontal cortex, whereas the EEG electrodes were placed over the motor cortex. This BMI processed 1) brain activity generated by mental arithmetic, which was detected from prefrontal recordings, and 2) motor commands generated by hand tapping, which were extracted from motor cortical activity.

Interestingly, NIRS methods can detect not only slow responses related to hemodynamics but also fast responses (with a millisecond temporal resolution) related to the scattering of light by neuronal membranes (315, 762, 861). This signal is weak, however, and has not been used for single-trial detection of neural activity.

D. Functional MRI

Functional MRI (fMRI) is another method for measuring brain hemodynamic responses (165, 276, 509, 772). fMRI-based BMIs derive their control signals from blood oxygen level dependent (BOLD) activity measured with an MRI scanner (744, 745, 847). fMRI has a relatively low temporal resolution (1–2 s), and the lag between neuronal activity and BOLD response is on the order of 3–6 s. The main advantage of fMRI-based systems is their spatial resolution that allows monitoring the entire activity of the brain. fMRI-based BMIs typically use slices with 5 mm thickness; each slice is represented by a $64 \times 64$ image with 3- to 4-mm voxels.

fMRI-based BMIs usually utilize visual feedback to display to subjects their own brain activity. For this purpose, various displays have been used, including functional brain maps (877), scrolling graphs of BOLD activity (139, 848), and virtual reality (745). The great advantage of fMRI is the ability to target a localized area as the source of a BMI signal. Weiskopf et al. (848) and Caria et al. (113) utilized imaging of the anterior cingulated cortex to develop a BMI for self-regulation of emotional processing. Additionally, BMIs have been developed for self-regulation of activity in the supplementary motor area (SMA) and Broca area (745).

Several studies have employed fMRI-based BMIs to control computer cursors, avatars, and robotic arms. In the study of Yoo et al. (879), subjects were asked to perform several mental tasks (sequential number subtraction, covert speech, and imagery of right or left hand clenching) to generate BOLD activity that drove a cursor through a two-dimensional maze. Yukiashu Kamitani and colleagues extracted fMRI representation of individual finger movements and drove a robotic hand with those signals, as reported in *Scientific American* (365). In Lee et al. (472), vertical and horizontal movements of a robotic arm were generated from fMRI signals. Cohen et al. (151) demonstrated control of a whole body human avatar in virtual reality by an fMRI-based BMI driven by motor imagery. Although control of prosthetic limbs from an fMRI scanner is not applicable to everyday use, such systems can be useful as rehabilitation tools. Additionally, fMRI-based BMIs can be potentially used as clinical devices for treating neurological conditions, such as stroke (185, 877), chronic pain (140), emotional disorders (113), and psychiatric disorders (430).

**IX. BMIs WITH ARTIFICIAL SENSATIONS**

A. Restoration of Sensations

Sensory BMIs enable the flow of information from the external world to be delivered back to the subject’s brain (57, 199, 466, 469, 583, 678). These systems strive to repair damage to sensory neural circuitry. In principle, sensory BMIs could interfere with different levels of neural sensory processing, from peripheral receptors to the spinal cord, brain stem nuclei, thalamus, cortical sensory areas, and cerebellum.

For the development of efficient sensory BMIs, it is important to understand that sensory processing does not depend only on a unidirectional flow of information from the peripheral receptors or sensory organs to hierarchically higher processing stages. Top-down modulatory signals (e.g., describing influences to the lower-order areas from the higher-order brain regions) are vital for sensory processing in awake subjects (303, 306, 454, 464, 617, 739). Such modulatory signals are essential during the execution of voluntary movements (126, 572, 714–716, 756, 760) and active sensory exploration of the environment (176, 314, 454, 615, 617).

Sensory impairments can take many forms, from a complete loss of sensation caused by destruction of peripheral receptors and nerves to impairments of certain aspects of sensory processing that occur when cortical or subcortical areas are damaged. For example, following extensive lesions to the primary visual cortex, patients (163, 769, 849) and monkeys (164) do not perceive visual stimuli, but may retain an ability to utilize visual information. This phenomenon, called blindsight, is mediated by subcortical visual structures like the superior colliculus. Additionally, damage to cortical areas of the so-called ventral visual stream produces deficits of visual object recognition, whereas damage to the dorsal stream areas impairs spatial visual processing and
visually guided movements (312, 313, 346). While such peculiar sensory disabilities could likely be treated with BMIs in the future, current implementations of sensory BMIs deal mostly with cases of damage to peripheral sensory receptors, sensory nerves, or spinal tracts. In these cases, there is a loss of normal sensation, but higher-order sensory areas remain intact and could still process sensory information if it is delivered to them using a BMI. Accordingly, sensory BMIs attempt to mitigate the devastating effects of peripheral lesions by linking these intact brain areas to artificial sensors.

We will focus on sensory BMIs that enable artificial tactile sensations. In such BMIs, electrical stimulation is usually used to reactivate sensory responses (260, 598, 676, 784). Additionally, several papers have recently employed optogenetic methods to induce somatosensory sensations (536, 875). Stimulation can be applied to somatosensory cortex (368, 598, 676, 784), thalamus (182, 321, 479, 605), and peripheral sensory nerves (183, 662, 683, 791).

Tactile sensations evoked by electrical stimulation of the surface of the postcentral cortex without eliciting movements were first described in 1909 by Harvey Cushing (178); they were later extensively studied by Wilder Penfield (623). Penfield’s patients most often reported sensations of numbness or tingling, rarely pain. The modern era in this research started with the experiments of Ranulfo Romo et al. (676) who employed small currents injected through a microelectrode, the method called ICMS, to evoke tactile sensations in monkeys. Romo’s monkeys started with learning a sensory discrimination task where they compared two vibrotactile stimuli applied to their hands one after another. The animal reported, by pressing a button with an opposite hand, which of the two vibrations had a higher frequency. Next, the first stimulus in the sequence remained vibrotactile, whereas the second one was an ICMS train applied to S1. The task was again to compare the frequencies at which the stimuli were presented. Surprisingly, monkeys began to successfully compare the vibrotactile and ICMS patterns with very little practice. This result suggested that sensations resembling skin vibration could be evoked artificially with ICMS of S1. Romo et al. (676) penetrated S1, with a microelectrode placed at a new location every day; they did not implant those microelectrodes. With this method, they could not study long-term changes in the ICMS-induced artificial sensations.

A long-term study of ICMS effects with implanted microelectrodes was conducted by our laboratory (260) [FIGURE 17]. The experiments were conducted in owl monkeys chronically implanted with cortical microelectrode arrays. The experimental task consisted of having animals reach and open one of two doors. Animals were searching for a piece of food that was hidden behind one of the doors. In each trial, the location of the food was cued by an ICMS train. Progressively, more complex ICMS patterns were employed as the animals learned novel tasks. Monkeys were first required to simply detect the presence of ICMS. Next, they had to discriminate temporal patterns of ICMS, and finally they discriminated spatiotemporal ICMS patterns delivered through multiple electrodes. Although it took several weeks for monkeys to learn the initial, simple task, their ability to interpret new ICMS patterns clearly improved after several months of training with ICMS. After this initial learning phase, animals could acquire a new and more difficult task in just a few days. This result indicated that ICMS progressively generated a new sense, some sort of artificial touch sensation, that monkeys could readily utilize. In fact, it seems that some degree of generalization was achieved by these monkeys after a few months of training, which allowed them to learn new tasks that involved ICMS faster than when they were naive in terms of experiencing ICMS. We also conducted a study in rhesus monkeys where ICMS of the S1 instructed the animals about the direction of joystick movement they had to produce in a trial (597). Like the owl monkey experiment, rhesus monkeys learned that task after several days of training.

Talbot et al. (784) asked if sensations from different skin locations could be evoked using ICMS. They applied ICMS through different microelectrodes of the arrays implanted in the hand representation of S1. Monkeys reported with eye movements where on the hand they felt the stimulus. Predictably, these experiments confirmed the well-known S1 somatotopic organization (394). Additionally, it was determined that monkeys could discriminate ICMS intensity and match it to the pressure applied to the hand using a mechanical probe. Virtually the same experiment was recently conducted in a tetraplegic patient implanted with a Utah array in S1 (262). The patient correctly matched cortical stimulation sites to different hand locations. Stimulation applied through ECoG grids has been shown to evoke somatosensory sensations, as well (589, 638). This stimulation method evoked sensations of tingling, numbness, and temperature. Electrical stimulation of both the precentral and postcentral locations was effective in producing these sensory effects.

Several studies explored peripheral nerve stimulation as a method to provide humans with artificial tactile sensations. One study (791) employed peripheral nerve cuff electrodes implanted in two patients with arm amputation for more than 1 yr. Patterned electrical stimulation of the nerves produced touch perceptions in the phantom hands that the patients described as being natural (tapping, pressure, moving touch, and vibration); the sensations changed with modifications in the stimulation pattern. These artificial sensations improved the subjects’ performance with a prosthetic hand. In the other study (183), phantom hand sensations were evoked in amputees using electrical stimulation of the median or ulnar nerve delivered through a 96-channel Utah
MicrostimulationTransparent barrier
Food Reward hidden behind one of two doors

A

Microstimulation Waveform

100 μsec
50-200 μA
50-200 μA
100 μsec

10 msec
x 3

E

150 msec
Basic Task
cues right
discriminated from
No Stimulation
cues left

G

No Stimulation
cues right
discriminated from
150 msec
100 msec
cues left

F

Spatiotemporal Discrimination

EP1
EP2
EP3
EP4
cues right
cues left
discriminated from
arrays, which remained implanted for 30 days. The same stimulation approach was utilized to reproduce sensations from a hand prosthesis that performed grasping tasks (662).

Several somatosensory BMIs have been demonstrated in rats, taking advantage of the exquisite tactile skills of these animals. A study of John Chapin’s laboratory (790) reported a BMI that guided rat navigation through three-dimensional structures. Steering cues were provided by ICMS of S1, whereas locomotion was reinforced by ICMS applied to the medial forebrain bundle, a structure known to be part of the brain’s reward system. A human operator used this BMI to steer rats over complex terrains. Venkatraman and Carmena (820) developed an active sensing paradigm based on ICMS of rat S1. ICMS was delivered when a whisker crossed a spatial location designated as a target. Rats were rewarded for localizing the invisible target and crossing it several times with the whisker.

Using rats as an experimental model, Thompson et al. (798) showed that ICMS of S1 could be used to substitute or augment the animal’s natural vision. Their BMI allowed rats to use the S1 cortex to perceive, or “touch,” otherwise invisible infrared light. Light from infrared (IR) sources was detected by head-mounted sensors and converted into ICMS applied to the rats’ S1 representation of their facial whiskers. Initially, rats took 4 wk to learn to use a BMI with a single infrared detector to find reward ports that emitted infrared light. An upgrade to this system included four IR sensors that provided a panoramic infrared vision (341). Using this system, a new group of rats took only 3 days on average to find the same infrared sources. After rats learned to utilize this BMI, electrophysiological recordings revealed that S1 neurons developed multimodal receptive fields that represented both somatosensory responses from the facial whiskers and infrared light generated in the animal’s surroundings. These results showed that, even in adult animals, primary cortical areas can incorporate new sensory representations, leading to the emergence of multiple and overlapping sensory maps simultaneously sustained by the same neuronal populations.

Sensory substitution through haptic stimulation of the subject’s body is an alternative to using the neurostimulation approach, which is particularly relevant to neurorehabilitation. A study by the Walk Again consortium (737) used haptic stimulation to restore the sensation of autonomous walking to paraplegic patients. For this purpose, a new paradigm was developed for reproducing lower limb somatosensory feedback in paraplegics by substituting sensations generated by a haptic display placed on patients’ forearms for the normal sensation generated by walking legs. Initially, leg movements were simulated by making an avatar of the patients move in an immersive virtual reality environment. Patients used goggles to observe their avatars moving on different ground surfaces while a haptic display was used to deliver a wave of tactile stimulation to the skin of their forearms. The use of this haptic display induced patients to experience the perception of walking on various surfaces: grass, a paved street, or beach sand. Moreover, patients perceived leg movements during the swing phase of the avatar legs and experienced the perception of their feet rolling on the floor, despite the fact that their legs were completely paralyzed. These results showed that virtual reality training combined with haptic stimulation resulted in the assimilation of the virtual lower limbs in the body representation present in the patients’ brains. These findings suggest that, in the future, the addition of rich haptic feedback to rehabilitation devices will be essential to restore realistic perceptual experience in paralyzed patients.

In addition to electrical stimulation, optogenetic stimulation has been steadily gaining popularity (875), so it is plausible that this method will be used in sensory BMIs in the future. Another stimulation method employs ultrasound (473, 674). Recently, implantable microcoils have been developed for magnetic stimulation (474).

**B. Brain-Machine-Brain Interface**

Brain-machine-brain interfaces (BMBIs), also called bidirectional BMIs, perform both the extraction of motor command signals from raw brain activity and the delivery of sensory feedback to the brain (57, 255, 469, 583) or peripheral nerves (543). This approach was pioneered by our laboratory (597, 598) (FIGURE 13). In our experiments, rhesus monkeys were chronically implanted with microelectrode arrays in M1 and S1. M1 implants were used for the extraction of motor commands, and ICMS was delivered through S1 implants. The motor loop of this BMBI controlled movements of an avatar arm shown on a computer screen placed in front of the animals. Monkeys used this avatar arm to actively explore a set of virtual objects (2 or 3 gray circles) rendered in the virtual space they searched. The objects were visually identical but differed in terms of their artificial

![FIGURE 17.](http://physrev.physiology.org/) Long-term experiments in owl monkeys on reaching movements cued by intracortical microstimulation of somatosensory cortex (S1). A: diagram of the experimental task. After a barrier was lifted, monkeys reached toward one of two doors; a food pellet was behind one of them. B: location of cortical implants. S1 implant was used to deliver microstimulation. C: microstimulation parameters. D–F: stimulation patterns. The first task (E) required reaching to the right if a sequence of microstimulation pulses was delivered. If no stimulation was applied, monkeys reached to the left. In the second task (G), the rule was reversed: monkeys reached leftward in response to microstimulation. The third task (D) employed two different temporal patterns of microstimulation. The fourth task (F) used spatiotemporal patterns of microstimulation produced using four pairs of implanted microwires. [Adapted from O’Doherty et al. (597).]
texture. Monkeys had to assess the objects’ texture by using the BMBI to scan their avatar hands over the surface of the virtual objects. When a monkey’s avatar hand came into virtual contact with the surface of a given object, a pattern of ICMS was delivered to the hand representation of the animal’s S1. Monkeys had to identify a specific virtual texture using this BMBI and then hold the avatar hand over it to obtain a fruit juice reward. The implementation of this BMBI had a caveat: since ICMS evoked electrical artifacts that occluded the neuronal spikes, recordings and stimulation could not be conducted simultaneously. This issue was solved by switching from recording to stimulating every 50 ms. Although this approach resulted in a loss of some neuronal data, the BMBI still performed well because several hundred neurons were recorded simultaneously. A similar BMBI was reported by Richard Andersen’s group (437). In that system, ICMS was applied to S1, whereas BMI control commands were extracted from PPC.

Several recent studies have implemented bidirectional interfaces with peripheral nerves. Davis et al. (183) demonstrated real-time control of a robotic finger by amputees using multielectrode recordings from the median or ulnar nerves. The decoding was performed by a Kalman filter. The same electrodes were used to deliver sensory feedback using electrical stimulation. The other study (662) reported a myoelectric interface that amputees could control using surface EMGs to produce grasping movements using a robotic hand. Grasp force feedback, produced by robotic sensors, was delivered using intracortical stimulation of the median and ulnar nerves; stimulation intensity was proportional to the sensor signal. This bidirectional setup enabled the subjects to maintain three force levels without looking at the robotic hand.

In addition to BMBIs that provide sensory feedback from an external actuator, several recent demonstrations of closed-loop activity-dependent stimulation can be described as BMBIs. In these systems, neuronal activity is recorded from a brain area and then converted into a pattern of electrical stimulation delivered to the same or a different area. Such feedback loops may serve different purposes. Andrew Jackson and his colleagues at Eberhard Fetz’s laboratory employed a neural implant to form and strengthen an artificial connection between two sites in the motor cortex of freely behaving monkeys (381, 382). The implant triggered electrical stimulation in one cortical location with neuronal discharges recorded from a different site. Several days of operation of this implant produced a stable cortical reorganization that was evident from the changes in wrist movements evoked by electrical stimulation applied to each site. Wrist movements evoked from the implant’s recording site started to resemble those evoked from the stimulation site, which indicated that a Hebbian potentiation of synaptic connections occurred for the artificial connection. The authors suggested that this approach could be used for neurorehabilitation in the future.

Lucas and Fetz (513) employed EMG-triggered cortical stimulation to induce a similar targeted reorganization of cortical motor output. They observed that the stimulated cortical site became associated with the activity of the recorded muscle, even though that particular muscle was not represented by neurons in that cortical location previously (513). Yet another study by the Fetz laboratory (592) demonstrated that spinal stimulation, triggered from cortical spikes, could modify the strength of corticospinal connections in a manner consistent with spike-timing-dependent plasticity.

Several studies have shown that closed-loop stimulation systems can lead to partial recovery of function in neurological conditions resulting from injury or disease. Guggenmos et al. (330) employed a functional bridge connecting motor and somatosensory areas of the rodent brain to promote recovery of motor skills after traumatic brain injury. McPherson et al. (540) used EMG-triggered spinal stimulation to facilitate recovery after spinal cord injury in rats. Overall, these studies showed that BMBIs could be used to plastically modify neural connectivity and promote functional recovery. Our laboratory developed a closed-loop stimulation system for epilepsy control (618). In that study, rats were treated with pentylenetetrazole to provoke epileptic seizures. The system detected the seizure episodes in cortical LFPs, and applied electrical stimulation to rat spinal cord to suppress the seizures. This approach reduced the frequency of seizure episodes by 44%. In the future, a similar approach may prove useful for the treatment for drug-resistant epilepsy.

A stimulation system has been suggested as a potential prosthetic system for improving memory (62, 63). In these studies, a multiple-input, multiple-output model reproduced the associations between CA3 and CA1 regions of the rat hippocampus. Neuronal ensemble recordings were conducted in CA3 and CA1 of rats performing a delayed-nonmatch-to-sample memory task. A nonlinear MIMO was trained to predict CA1 activity based on CA3 patterns. The predicted patterns of activity were then delivered to CA1, using electrical stimulation through the same electrodes that recorded neuronal spikes. The stimulation improved task performance in normal rats and restored performance in rats with a pharmacological block of hippocampal synaptic transmission. The authors suggested that this approach could be used to restore long-term memory function in patients with damage to hippocampus and its interconnected structures.

X. COGNITIVE BMIs

Cognitive BMIs or cognitive neural prostheses deal with brain activity related to higher-order functions, as opposed
to more simple motor and sensory functions (625, 792). Although the distinction between higher-order and lower-order functions is not clear cut, by convention BMIs are called cognitive if they work in the domains of cognitive states (348, 886), executive functions (349), decision making (21, 335, 343), memory (62, 63), attention (35, 417), and language (98, 99, 327, 482, 834, 839).

Several intracranial cognitive BMIs have been developed, mostly dealing with decoding of different aspects of motor decisions during the periods of immobility preceding movement onsets. For example, Hasegawa et al. (343) decoded go versus no-go decisions, and the prepared saccade direction from the activity of monkey superior colliculus neurons (343). Musallam et al. (564) decoded the representation of expected rewards and motor decision from the neural activity recorded in the cortical parietal reach region. In that study monkeys were engaged in an instructed-delay task where they prepared an arm movement, but withheld it for several seconds. In the beginning of each behavioral trial, the monkey was shown a cue that indicated what kind of reward would be given. A Bayesian algorithm was applied to decode expected reward and target location simultaneously.

In our laboratory, a BMI approach was employed to extract decisions involved in reprogramming a motor goal (376). Monkeys performed a center-out task, where they moved a cursor towards screen targets using a joystick. Neuronal ensemble activity was recorded from M1 and S1 arm representations. Monkeys started the trials by placing the cursor at the screen center. Next, a target appeared at an off-center location. In some trials, that initial target appeared for 50–250 ms, and then it was replaced by a new target at a different screen location. We found that both the emergence of the decision to move to the initial target and the new decision to cancel that motor plan and move to a new target could be decoded from population M1 activity. Target locations were decoded using an LDA classifier. This analysis showed that M1 activity initially represented the first target, then simultaneously represented both targets, and eventually shifted to represent the new target only. Based on these findings, we proposed that such decoding of covert motor planning could improve motor BMIs by equipping them with the capacity to detect motor decisions early and inhibiting them if the user decides to cancel a prepared action or chooses a different one.

In the other study, we decoded representation of time from M1 and PMd activity in the absence of overt behavior (468). Monkeys were trained to perform self-timed button presses, where they touched a button with their hands, maintained contact for 3–4 s, and then released the button. We found that, while monkeys self-timed the required interval and did not produce any movements, their M1 neurons exhibited ramping activity patterns. We then employed a Wiener filter to decode the representation of temporal intervals from these M1 patterns. Such decoding of action timing could be useful for developing BMIs that enact typical motor behaviors where movements are intermingled with periods of immobility.

Motor imagery, widely used in noninvasive BMI (5, 28, 272, 551, 773, 840), can be considered a cognitive component of a BMI. Richard Anderson’s group recently decoded motor-imagery from the intracranial PPC recordings in a tetraplegic human (3). PPC is engaged in higher-order aspects of motor behaviors (18). In the tetraplegic subject, motor imagery clearly activated different PPC neurons, depending on the specific action being imagined (e.g., imagining hand movement to the mouth or ear, imagining shoulder rotation, etc.). Additionally, PPC neurons responded to the imagery of movement goal, movement trajectory, and the type of movement. All these variables were successfully decoded from the activity of PPC neuronal populations. Moreover, the subject learned to control a robotic arm by imagining movements.

As intracranial recording methods become more routine in clinical studies, we will probably see a rapid development in BMIs related to human cognitive processes. Advances in this new domain will likely contribute to the emergence of new clinical applications for BMIs, as well as the incorporation of fundamental knowledge about the neurophysiological involved in higher brain functions.

**XI. BRAIN-TO-BRAIN INTERFACES AND BRAINETS**

The growth of BMI research gave rise to a large variety of spin off experimental paradigms. In one of the variations of the classical BMI approach (FIGURE 18), multiple animal (or human) brains can be connected to each other to establish a direct brain-to-brain communication linkage, called brain-to-brain interface (BTBI) (616). In the other variation, several individual brains collaborate on a common motor task by establishing a network of brains, or a Brainet (614, 657) (FIGURE 19).

By definition, BTBIs allow multiple animals to exchange information using a protocol that incorporates both neural recording and stimulation. The pioneering BTBI was implemented in rats (616). In that study, the first animal performed the role of information encoder, and the second animal was the decoder of a simple binary message. The binary message represented the encoder rat performing a two-choice behavioral task (active tactile discrimination or responses to a visual stimulus). The encoder rat’s neuronal firing rates, recorded from either the S1 or M1, depending on whether the rat performed a tactile discrimination or a visuomotor task, underwent a sigmoid transform and then were converted into patterns of ICMS
applied to S1 or M1 of the decoder rat. This latter animal could be located next to or far apart from the encoder. On average, the decoder reproduced the behavioral choices of the encoder rat in about 70% of the trials. During operation of the BTBI, the encoder rat received an additional reinforcement. Pais-Vieira and colleagues noticed that following an error by the decoder rat, the encoder rat adapted both its behavior and cortical activity to generate cleaner neuronal signals to be broadcast to its partner. Invariably, the decoder rat performed better after this encoder’s adaption.

In the next study by Pais-Vieira et al. (614), several rat brains were connected to a network of brains - named a Brainet - that performed several elementary computations, like discrimination of ICMS patterns by several rats simultaneously to improve overall discrimination accuracy, or retaining information in their collective memory by transferring it from rat to rat (614). ICMS served as an input to such a Brainet, while the output was derived from cortical activity of the participating animals. Essentially, the Brainet acted as an organic computer that processed input data through a network of living brains.

These initial publications were followed by a number of studies by different groups unified by a common theme of connecting of the brains of different organisms. Yoo et al. (878) connected the brain of a human to the spinal cord of a rat. The human operated an SSVEP-based BMI to generate “go” commands to an anesthetized rat. The command was executed by applying transcranial focused ultrasound to the rat motor cortex, causing the movement of the animal’s tail (878). In another study BTBI connected two different species (488). The human attempted to make the cockroach walk along an S-shape track, and succeeded in 20% of cases. In the other study (727), the premotor cortex of one monkey was linked to the spinal cord of a second, anesthetized primate. The second monkey’s hand was attached to a joystick. The first monkey generated motor intention commands while looking at a computer screen that showed a cursor and a target of movement. This intention command was extracted from premotor cortex activity and translated into a stimulation pattern applied to the spinal cord of the second monkey, causing the joystick movement, which in turn moved the cursor on the first monkey’s screen. A proof of concept study (265) showed that gene expression can be controlled by brain signals. In that study, a human operating an EEG-based BMI optogenetically controlled the expression of designer cells. The designer cells were either in culture or in subcutaneous implants in mice.

Several BTBIs have been demonstrated in humans. Grau et al. (316) had one human subject, the emitter, operate a motor-imagery EEG-based BMI. The binary output of that BMI was delivered to the brain of the second subject, the receiver, using TMS pulses applied to the visual cortex. Depending on whether the transmitted signal was “1” or “0,” a robot placed the TMS coil over the area where stimulation induced or did not induce a conscious perception of phosphenes. Information transfer rates of 3 and 2 bits per minute were achieved for the BMI performance and message transmission, respectively. Rao et al. (661) employed a very similar BTBI design, with the dif-
ference that TMS was applied to the motor cortex of the second subject. Accordingly, the second subject responded with a TMS-induced hand movement that produced a touchpad press.

While the studies reviewed above emphasized direct communication between different brains, our laboratory recently demonstrated several Brainets that emphasized cooperation of multiple subjects to achieve a common motor goal of a typical upper limb BMI (657). In that study, two or three monkeys shared control of the movements of an avatar arm using their combined cortical activity. Three Brainet designs were tested. The first design, called shared-control Brainet, merged the outputs of two monkey brains. Cortical activity of each monkey was processed by a separate decoder. The decoder outputs were then averaged to set the coordinates of the avatar arm. Performance improvement was achieved because the averaging of contributions from both monkeys enhanced the signal and suppressed the noise. In the second design, called partitioned control Brainet, two monkeys performed together again, but they had different tasks. The first monkey generated neural control commands to move the avatar arm in the horizontal dimension, while the other monkey controlled the vertical dimension. In that Brainet, performance improved because each monkey made fewer errors in the simple, one-dimensional task. In the third Brainet design, named a triad Brainet, three monkeys cooperatively controlled three-dimensional movements of the avatar arm. Yet, each monkey performed a two-dimensional task, and all three together controlled three-dimensional movements of the avatar arm. [From Ramakrishnan et al. (657).]

**FIGURE 19.** Monkey brainet. A: diagram of the experimental setup. Up to three monkeys were seated in monkey chairs in separate rooms. Each monkey faced a computer screen that displayed a virtual avatar arm. The behavioral task consisted of reaching screen targets with the avatar arm. The avatar arm was controlled jointly by several monkeys. B: shared control task, where each of two participating monkeys contributed 50% to the (X,Y) position of the virtual arm. Locations of microelectrode arrays are shown below the task diagram. C: partitioned control task, where one monkey contributed to X position of the avatar arm, whereas the other monkey contributed to Y position. D: a 3-monkey task. Each monkey performed a two-dimensional task, and all three together controlled three-dimensional movements of the avatar arm. [From Ramakrishnan et al. (657).]
Somewhat similar cooperative systems have been developed using EEG-based controls by several humans. These include a BMI for spacecraft navigation controlled by two users (644), BMIs for group decision making (217, 645, 883), and a cooperative BMI for movement planning (841).

XII. BMI AS A POTENTIAL NEUROREHABILITATION THERAPY

Since the late 1990s, when BMI research began in earnest, the field has focused primarily on achieving two major goals: 1) to establish a new paradigm to investigate the dynamic physiological properties of distributed neural circuits in behaving animals, and 2) to explore the possibility of creating new assisted technologies, aimed at restoring upper, lower, or full body mobility in severely paralyzed patients. For the past decade, the focus on developing clinical applications based on BMIs has increased markedly, as noted throughout this review. Yet, no one had anticipated that this paradigm could provide benefits beyond the commonly stated goal of assisting patients in regaining mobility through the employment of a new generation of brain-controlled prosthetic or orthotic devices. Thanks to recent clinical studies, however, a third potential future application of this paradigm has been introduced: the use of BMIs as a neurorehabilitation tool (26, 27, 82, 204, 205, 737, 741, 751, 752, 819).

To date, noninvasive BMIs have been used as neurorehabilitation tools primarily in clinical studies focused on stroke victims. The main assumption motivating these studies has been that practice with a BMI that mimics movements of a paralyzed limb could facilitate brain plasticity and contribute to some level of motor recovery. For example, stroke patients can learn to operate an MEG-based BMI by modulating their $\mu$ rhythm recorded in the hemisphere ipsilateral to the lesion (104). In this study, the BMI opened and closed an orthosis that was attached to the paralyzed hand. This learning did not cause noticeable clinical improvements. However, long-term BMI training combined with physical therapy resulted in clear motor recovery (93, 659). As shown by the analysis of motor evoked potentials (MEPs), the recovery was related to enhanced neuronal activity in the hemisphere ipsilateral to the stroke site (87). Similar results were demonstrated by a study that combined a BMI-controlled robot with robot-assisted physical therapy (27, 30). Combining BMI training with virtual reality resulted in clinical improvements as well (64). Additionally, a combination of BMI control with transcranial direct current stimulation (tDCS) showed positive clinical results (753).

Much less research has been conducted on the effectiveness of BMI training in patients with SCI. In the first long-term study of this kind, Donati et al. (205) conducted BMI training of eight chronic paraplegic patients in a multi-stage rehabilitation paradigm, aimed at restoring bipedal locomotion through robotic lower limb orthoses. The core of this paradigm was based on the utilization of an EEG-based BMI that allowed patients to control multiple actuators, ranging from avatar bodies to two types of robotic walkers: a commercially available gait robotic system (Lokomat) (384) and a custom-designed lower limb exoskeleton. In addition to the traditional visual feedback, this BMI was also coupled with a haptic display system that delivered continuous streams of tactile information to the skin of the patients’ forearm. These artificial tactile/proprioceptive sig-

![Figure 20](http://physrev.physiology.org/) Sensory improvement after neurorehabilitation training. A: average sensory improvement (mean ± SE over all patients) after 10 mo training. B: example of improvement on Zone of Partial Preservation for sensory evaluation for two patients. [From Donati et al. (205).]
nals were generated either when the avatar body walked on a virtual surface, or when the patients walked with the help of the robotic devices. In the latter case, pressure sensors applied to the plantar surface of the robotic feet were responsible for generating signals depicting the feet’s contact with the ground during bipedal walking. Closing the control loop with this haptic display led patients to experience vivid lower limb phantom sensations, which included the illusion of experiencing leg movements even when they were operating the avatar body while remaining immobile themselves. Moreover, using the information delivered by the haptic display applied to the skin surface of their forearms, six out of eight patients could discriminate between three different types of surface in which the avatar body walked (e.g., sand, grass, and asphalt).

Despite being completely paraplegic, immobile from the level of the spinal cord lesion down, lesions ranging from (T4-T11), since the day of their spinal cord lesions (3–13 years earlier), and lacking any somatic sensation below the

**FIGURE 21.** Lower limb motor recovery. A: details of the EMG recording procedure in SCI patients. A1: raw EMG for the right gluteus maximus muscle for patient P1 is shown at the top of the topmost graph. The lower part of this graph depicts the envelope of the raw EMG, after the signal was rectified and low pass filtered at 3 Hz. Gray shaded areas represent periods where the patient was instructed to move the right leg, while the blue shaded areas indicate periods of left leg movement. Red areas indicate periods where patients were instructed to relax both legs. A2: all trials over one session were averaged (mean ± SD envelopes are shown) and plotted as a function of instruction type (gray envelope = contract right leg; blue = contract left leg; red = relax both legs). A3: below the averaged EMG record, light green bars indicate instances in which the voluntary muscle contraction (right leg) was significantly different (t-test, \( P < 0.01 \)) than the baseline (periods where she/he was instructed to relax both legs). Dark green bars depict periods in which there was a significant difference (\( P < 0.01 \)) between muscle contraction in the right versus the left leg. B: EMG envelopes and t-tests for all recording sessions, involving four muscles, for all eight patients: left and right gluteus maximus (GMx) and rectus femoris proximal (RFP) muscles. Color convention and figure organization follows the one of A. Data were collected after 7 mo of training for all patients and for all but patients P2 and P8 after 12 mo. [From Donati et al. (205).]
level of the lesion, after a 12 mo period of training with this BMI paradigm, all patients exhibited a very significant partial neurological recovery, which was characterized by the following: 1) an average expansion of 5 dermatomes, in pinprick, nociceptive sensation, in the zone of partial preservation,\(^1\) (below the level of the lesion) (FIGURE 20, A AND B); 2) an average 1–2 dermatome expansion in fine touch (FIGURE 20A); 3) significant improvement in proprioception and vibration perception below the level of the lesion; 4) recuperation of voluntary control of multiple muscles below the level of the SCI lesion, as measured by EMG recordings and direct force measurements. In some cases, patients regained the ability to produce multi-joint leg movements (FIGURE 21, A AND B); 5) marked improvement in the walking index; 6) an improvement in thoracic lumbar control (FIGURE 22B); and 7) restoration of peristaltic and bowel movements, bladder control, and improvement cardiovascular function (FIGURE 22C).

Because of this substantial neurological recovery, 50% of the eight patients were upgraded from a complete paraplegia (ASIA A \(n = 7\), ASIA B \(n = 1\)) to a partial paraplegia classification (ASIA C) at the end of 12 mo of training with this BMI-based protocol (FIGURE 22A). Longitudinal analysis of EEG recordings obtained from these patients during the 12-mo training period revealed that this partial sensory, motor, and visceral recovery was paralleled by an expan-

\(^1\)The area of the body, measured in dermatomes and myotomes, below the level of the spinal cord lesion, in patients classified as having a clinically complete lesion, that remained partially innervated.
sion of the representation of lower limbs in their primary sensorimotor cortex. Based on these results, Donati et al. (205) proposed that a combination of cortical and spinal cord plasticity, triggered by chronic use of a BMI that provided rich visuo-tactile feedback, may have rekindled remaining axons that survived the original spinal cord injury.

Overall, the study by Donati et al. (205) raises the concrete possibility that the future goals of BMI research may include the possibility of creating therapeutic procedures aimed at inducing some degree of neurological recovery in patients suffering from incomplete SCIs. As such, BMIs may become a true neurorehabilitation paradigm for these patients, instead of a mere assistive technology.

XIII. CONCLUSION

After a decade and a half of intense development, BMI research is currently witnessing a very rapid growth towards a broad range of potential clinical applications. This trend was originally driven by the expectation that BMIs may provide fundamental assistive tools for people who suffer from motor and/or sensory deficits. Recently, this expectation has been upgraded to reflect the possibility that BMIs may also become a new neurorehabilitation therapy that takes advantage of the phenomenon of brain plasticity to induce partial neurological recovery in severely disabled patients.

ACKNOWLEDGMENTS

Address for reprint requests and other correspondence: Miguel A. Nicolelis, Box 103905, Duke University, Durham, NC 27710 (e-mail: nicoleli@neuro.duke.edu).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

REFERENCES


