Theory of Nano-Electron-Fluidic Logic (NFL): A New Digital "Electronics" Concept

Héctor J. De Los Santos, *Fellow, IEEE*

***Abstract*—A new digital "electronics" concept is introduced. The concept, called nano-electron-fluidic logic (NFL), is based on the generation, propagation, and manipulation of plasmons in a 2-D electron gas behaving as an electron fluid. NFL gates are projected to exhibit femtojoule power dissipations and femtosecond switching speeds at finite temperatures. NFL represents a paradigm shift in digital technology and is poised as a strong candidate for "beyond-CMOS" digital logic.**

***Index Terms*—Beyond-CMOS digital logic, digital circuits, electron fluid, logic gates, surface plasmons, 2-D electron gas.**

I. Introduction

THE GROWTH witnessed by the semiconductor industry over the last 40 years was predicted by Gordon E. Moore in 1965, when he postulated that the number of components per chip would double every 18 months [1]. The exponential growth in circuit complexity, known as Moore's Law, translates into the economic argument that the cost of delivering digital functions on silicon wafers can be halved every two years. This cost reduction, in turn, has since been fueling the continuous expansion of the microelectronics industry with double-digit growth.

The question of identifying the challenges to continuing this exponential growth is addressed by a worldwide forum that includes semiconductor manufacturers, equipment and material manufacturers and suppliers, institutes, and universities and is published in the International Roadmap for Semiconductors (ITRS). The world's agenda for semiconductor technology developments over the next 15 years is captured via annual updates of the ITRS. These updates to the ITRS define the requirements for future generations of semiconductor chips as indexed by the minimum field-effect transistor (FET) gate length, e.g., 130, 110, 90, 65, 45, 35, 22 nm, etc.

The demise of Moore's Law, which began to manifest itself at FET gate lengths of the order of 100 nm [2], signaled that a diversification toward emerging, "beyond" CMOS, nanoscale devices was necessary to keep the industry growth pattern. In particular, the increase in power consumption due to increased leakage currents, short channel effects, source-drain tunneling and p-n junction tunneling [2], without a commensurate increase in operating speed, has meant that the point of diminishing returns for CMOS scaling is imminent. Examination of the factors causing this inefficiency finds that it is rooted in the convergence of a number of issues. On the one hand, there is the decrease in channel mobility and the increase in the interconnection resistance concomitant with the smaller process geometries, together with the increasing significance of defects. And on the other hand, there is the high level of complexity in both lithography and design, which have resulted in manufacturing costs rising to prohibitive levels. Accordingly, new device concepts and technologies that will shift the CMOS scaling paradigm and overcome the aforementioned limitations are beginning to be vigorously pursued.

In this paper, we advance the nano-electron-fluidic logic (NFL) concept. NFL's speed does not rely on particle transit time; hence, it is not affected by channel mobility limitations. Rather, NFL operates by steering the propagation *direction* of a surface plasma *wave* (SPW) set up in an electron fluid (EF) [3], [4], the EF acting as a wave guiding structure to the SPW (see Appendix A for an introduction to SPWs). The steering action is affected by *scattering* a "bias" SPW with a "control" SPW. In this context, device speed is mainly a function of SPW propagation velocity, which implies terahertz (THz) switching frequencies since SPWs propagate with a velocity two orders of magnitude greater than electrons [3], [4]. Furthermore, implementation of the NFL concept is totally compatible with available lithographic capabilities, hence, taking full advantage of established semiconductor-manufacturing infrastructure

II. Phenomenological NFL Description

A. Qualitative Discussion of Device Operation

Conceptually, the device operates as described in Fig. 1(a) (from top to bottom):

1) a *2-D electron gas* (2DEG) *behaving as a fluid*, i.e., an EF, is induced under a patterned gate;
2) a "bias" SPW is launched into the 2DEG EF;
3) a "control" SPW is launched at an angle to the bias SPW;
4) the two SPWs collide, resulting in the scattering of the bias SPW into a direction different from its prescattering direction;
5) the scattered SPW is detected. Fig. 1(b) depicts the sketch of a prototypical device that implements the sequence of events described in Fig. 1(a).

The prototypical NFL device has a patterned gate, terminals *source bias*, *source C1*, and *source C2*, at which SPWs may be launched, and terminals *drain O1* and *drain O2*, at which SPWs may be detected. A logic operation may be established as follows. Upon inducing the 2DEG EF under the patterned gate, a "bias" SPW is launched into terminal Source Bias. This

Manuscript received August 20, 2008; revised November 24, 2008 and June 17, 2009; accepted August 17, 2009. Date of publication September 9, 2009; date of current version July 9, 2010. The review of this paper was arranged by Associate Editor Dr. L. Dong.

The author is with the NanoMEMS Research, LLC, Irvine, CA 92623 USA (e-mail: hjd@nanomems-research.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNANO.2009.2031469

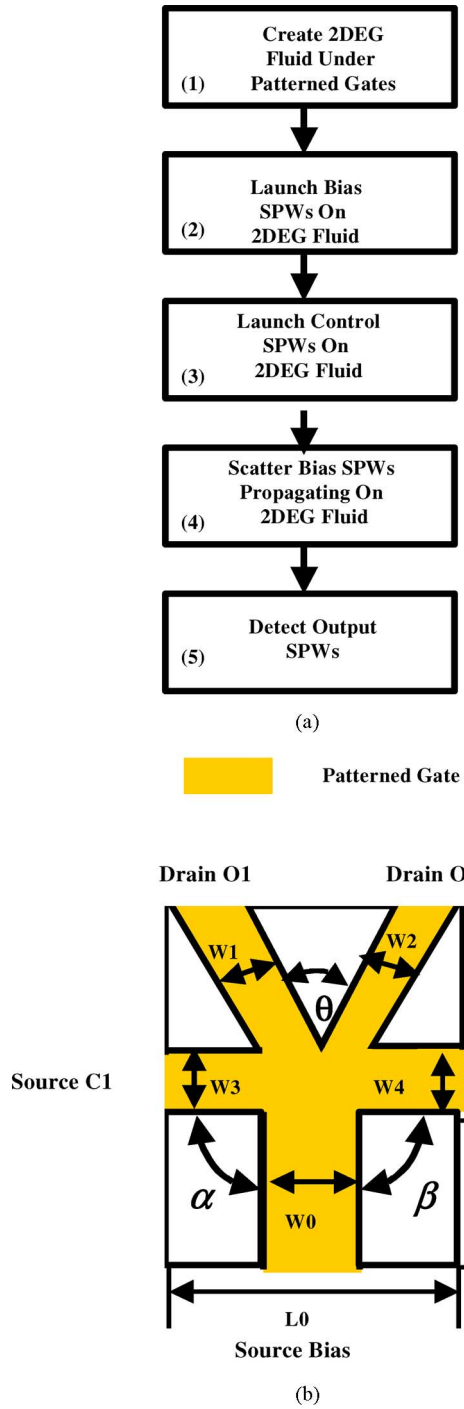

Fig. 1. (a) Conceptual sequence of steps describing NFL operation. (b) Top view of prototypical MOS implementation.

bias SPW propagates toward the drain O1 and drain O2 terminals. In the absence of control SPWs, which could otherwise be launched from either the source C1 or source C2 terminal, the bias SPW would be split so that (substantially) equal portions of it would detected at the drain O1 and the drain O2 terminals. If, e.g., the control SPW is launched from the source C2 terminal, but not from the source C1 terminal, then, upon reaching the junction, in the center of the structure, the bias SPW will be

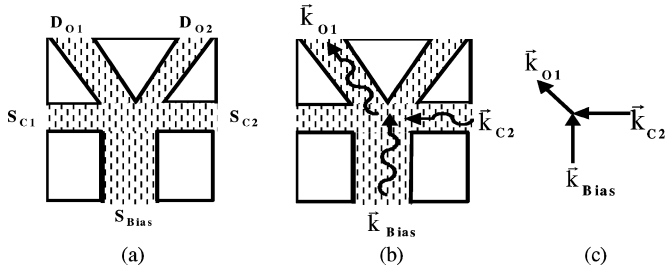

Fig. 2. (a) NFL FF device sketch showing source terminals $S_{\rm Bias}$, S_{C1}, S_{C2}, and (dashed) induced 2DEG EF underneath patterned gate. (b) SPWs with momentum vectors $k_{\rm Bias}$ and k_{C2} are *excited/launched* into terminals $S_{\rm Bias}$ and S_{C2} and propagate on surface of 2DEG EF. SPW with momentum $k_{\rm Bias}$ is diverted by SPW with momentum k_{C2} in direction of output SPW k_{O1}. (c) Momentum conservation determines k_{O1}.

TABLE I
FF TRUTH TABLE

Present State					Next State	
C1	C2	O1	O2		O1	O2
0	0	0	1		0	1
0	0	1	0		1	0
0	1	0	1		1	0
0	1	1	0		1	0
1	0	0	1		0	1
1	0	1	0		0	1

scattered/steered so it exits through the drain O1 terminal where it may be detected, see Fig. 2.

If the presence of an SPW is identified as a logic "1," and its absence as a logic "0," then the structure in Fig. 1(b) may be used to realize the truth table in Table I, which is that of a logic flip-flop (FF) [5]. Thus, SPW manipulation, as described earlier, realizes a new digital "electronics" concept.

The physical basis for SPW manipulation is now explained *qualitatively* with respect to Fig. 2, which uses dashed channels to represent 2DEG EF SPW *waveguide*, Fig. 2(a), and wavy arrows to represent SPWs propagating on the 2DEG EF, Fig. 2(b). The patterned gate is NOT, as shown in Fig. 2, only what goes on underneath it in the 2DEG EF.

In this NFL implementation, Fig. 2, SPWs with momentum vectors $\vec{k}_{\rm Bias}$ and $\vec{k}_{C2}$ are *excited/launched* into terminals $\mathbf{S_{Bias}}$ and $\mathbf{S}_{\rm C2}$ and propagate on the 2DEG EF. Upon collision at the junction/center of the structure, the SPW with momentum $\vec{k}_{\rm Bias}$ is scattered (diverted/steered) by SPW with momentum $\vec{k}_{C2}$ in the direction of output SPW $\vec{k}_{O1}$. The presence of SPW $\vec{k}_{O1}$ at $\mathbf{D_{O1}}$ is detected as a logic "1." The absence of an SPW at $\mathbf{D_{O2}}$ is detected as a logic "0." As shown in Fig. 2(c), conservation of SPW momentum determines $\vec{k}_{O1}$. By controlling the launching of SPWs, as per the truth table in Table I, the FF logic function may be realized.

The *innovation* of the concept is the following.

1) Switching is effected by steering the propagation direction of an SWP by scattering it with another SPW. For example, bias SPW $k_{\rm Bias}$ may be caused to be scattered into output SPW k_{O1} when impinged upon by control SPW k_{C2}.

2) The presence of SPW k_{O1} represents, e.g., a logic "1."
3) Due to nanometer scale dimensions and SPW propagation velocity $\sim 10^8$ cm/s [4], switching times of the order of femtoseconds, or switching frequencies $\sim$THz, can be obtained.
4) Since an SPW can be excited by any nonzero dc current, and its velocity is approximately independent from current, negligible power consumption is required to maintain the proper conditions for their generation, namely, the existence of the 2DEG EF. Power consumption will be set by the minimum detectable current. NFL is, thus, the basis of a radically new "beyond-CMOS" logic technology.

III. Technical Discussion of Device Operation

A. Origin of the Concept

The concept originates from the realization that SPWs propagating in a gated 2DEG structure may be *guided* so that a first SPW, when caused to collide with a second SPW, may steer/deflect it into a desired direction, in which it can be detected. The fundamental theory supporting the concept rests on two phenomena: 1) plasma wave generation and propagation in a *sheet* of very high carrier density; 2) the physics of plasmon–plasmon scattering.

In this Section, we summarize the fundamental physical theory, which supports this idea and which is behind our NFL device.

B. SPW Propagation in Sheet of Very High Carrier Density

The SPWs in a gated 2DEG in the inversion layer of a FET were first studied theoretically by Nakayama [4]. He found that the SPW dispersion relation obeys $\omega = s_L k$ and $\omega = s_H \sqrt{k}$, at frequencies much lower and much greater, respectively, than the plasma frequency of the free carriers $\omega_0^2 = n_s e^2 / m^* c$, where n_s is the free-carrier density, $m*$ is the electron effective mass, c is the speed of light, and k is the wave vector. s_L and s_H are the respective speeds of SPW propagation at low (L) and high (H) frequencies.

Experimentally, the existence of SPWs was first reported by Allen *et al*. [6], who detected the 2-D SPW in a silicon inversion layer, via the observation of infrared absorption, and by Tsui *et al*. [7], who observed weak infrared emission.

C. SPW Generation in Sheet of Very High Carrier Density

A method to generate SPWs in the inversion layer of an FET was first proposed by Dyakonov and Shur [3]. The required conditions were given as: 1) a large gate–source bias U_0 so as to produce a high enough carrier density in which electron–electron collisions would dominate and induce drift velocity-limited transport and 2) a constant drain current setting up an average source–drain electron velocity v_0 and a corresponding electron flux per unit width $j = n_s v_0 = C U_0 v_0 / e$ [3]. Here C is the gate capacitance per unit area.

In the *absence* of drain current, application of U_0 *disturbs* the high-density charge sheet equilibrium [3] and results in the creation of an SPW with linear dispersion, which propagates with a wave velocity $s_L = \sqrt{eU_0/m^*}$ [3]. Thus, the high-density carrier sheet would act as an SPW *waveguide*. The maximum length of propagation, however, is limited by two main decay mechanisms, namely, electron scattering by phonons or impurities, captured by the momentum relaxation time τ_p, and the viscosity of the electron fluid $\nu = v_F \lambda_{ee}$, where v_F is the Fermi velocity and λ_{ee} is the average distance between electrons. The viscosity gives rise to a damping time constant $\tau_v = vk^2$. The ability of an SPW of velocity $v_{\rm SPW}$ to propagate a distance L, then, depends on whether the conditions $L/v_{\rm SPW} < 1/2\tau_p$ and $L/v_{\rm SPW} < \tau_v$ are satisfied. In other words, it depends on whether "collection" of the SPW can be effected before it decays. To ensure the adequate longevity of the SPW, its velocity must exceed a certain threshold velocity $v_{\rm Th}$ determined by both scattering and viscosity and captured by an effective meantime between collisions τ.

If their propagation length were sufficiently long, SPWs could be employed for signal processing functions, the only energy required being that to launch them. The energy to launch an SPW may be estimated as follows. We visualize the plasma oscillation as resulting from the motion of the electron density in the 2DEG, of total mass m, with respect to the positive background charge, and its motion under the electrostatic restoring force [8]. This, being analogous to a spring-mass system [9], allows us to estimate the maximum stored energy, as $E_s = m v_{\rm sat}^2/2$, where $v_{\rm sat}$ is the saturation velocity of electrons. Then, the energy dissipated (i.e., spent in setting the plasmon into motion) is given by $E_d = E_s/Q$, where Q is the quality factor of the system. As with any other oscillating system, the Q of the 2DEG is given by $Q = \omega_0/2\alpha = \omega_0 R_{\rm 2DEG}/2L_{\rm 2DEG}$, where ω_0 is the plasmon frequency, α is the damping constant, and $R_{\rm 2DEG} = m^*/ne^2 A\tau$ and $L_{\rm 2DEG} = m^*/ne^2 A$ are the 2DEG resistance and *kinetic inductance*, respectively [10]. The energy cost (dissipated) to launch a plasmon, then, is given by $E_d = m v_{\rm sat}^2/2\omega_0\tau$.

The plasma frequency ω_0 is in general a function of the geometry. For a 2DEG, in particular, Dyakonov and Shur [11], [12] obtained it as $\omega_{0\_2\rm DEG} = \sqrt{e^2 n_s k/2\varepsilon_r \varepsilon_0 m^*}$, where ε_0 is the permittivity of vacuum, ε_r is the relative dielectric constant of the semiconductor, and k is a specific wave vector, which depends on the 2DEG geometry and boundary conditions [12]. For an *ungated* 2DEG, with grounded source and open drain, $k = \pi(2l-1)/2L$, where L is the source–drain distance and l is an integer standing for the mode index [12]. For a gated 2DEG with gate–source bias U_0, $k \cong \sqrt{2\pi n_s}$ [11]. The corresponding plasmon propagation velocity, is given by $s_{2D} = \sqrt{e^2 n_s d/m^* \varepsilon_r \varepsilon_0}$ [13], where d is the gate-to-channel separation.

The energy required to launch a plasmon and their subsequent propagation speed is now computed. Using the aforementioned expressions, and assuming $\varepsilon_r = 11.8$, $m^* = 1.08 m_0$, $n_s = 6.295 \times 10^{16}$ m^{-2}, $v_{\rm sat} = 2 \times 10^5$ m/s, $\tau = 10^{-13}$ s, a gate length $L = 140$ nm, and width $W = 50$ nm, and a gate-channel separation $d = 1.2$ nm, one obtains $E_d = 1.143 \times 10^{-3}$ fJ, and $s_{2D} = 1.373 \times 10^7$ m/s. Thus, the SPW would propagate a distance of 140 nm in 10 fs. Signal processing functions based on SPWs of long-propagation lengths would be an extremely

low-power dissipation and high-speed technology. This is what is at the heart of the digital logic concept proposed, Fig. 1(b).

The picture of SPW transport has been experimentally verified by exciting the SPWs with electromagnetic waves of frequency ω in gated 2DEGs of varying lengths. In particular, the measured drain-to-source voltage generated as a photoresponse for a device with $L = 30$ nm was 10 mV greater than that for a device with $L = 50$ nm, when illuminated with 119 GHz radiation at 300 K [14].

In the *presence* of drain current, the SPW dispersion relation becomes $k = \omega/(v_0 \pm s_{\rm SPW})$, where the plus sign pertains to propagation along the direction of the current, and the minus sign to propagation in the opposite direction [3]. In this context, Dyakonov and Shur [3] have shown that if, in addition to the fixed voltage U_0 at gate–source end of the channel, a fixed current is set at the drain end, then the SPW may grow as it bounces back and forth from the ends of the 2DEG channel cavity. This growth occurs because the amplitude ratio of the reflected and incoming wave, given by $(s_{\rm SPW} + v_0)/(s_{\rm SPW} - v_0)$, grows when the electron velocity is less than the SPW velocity, i.e., $v_0 < s_{\rm SPW}$, but greater than the threshold velocity $v_{\rm Th}$ [3]. Under these conditions, electromagnetic radiation of a frequency $\omega_0 = \pi s_{\rm SPW}/2L$ is generated. Recently, Dyakonova *et al*. [15] reported the room-temperature generation of THz radiation in 50 nm gate-length InAlAs/InGaAs high-electron mobility transistors. The generation and propagation of SPWs has, thus, been established [3], [14]–[16], and we can identify two clear regimes of application, namely, THz signal detection and generation and SPW propagation and manipulation for signal processing. The latter is the new concept we propose.

Since the plasmon velocity in the direction of current flow is increased, and that in the opposite direction is reduced, one can see that setting up a drain current facilitates achieving the condition for extending the propagation length of the SPW, and favoring the direction of current flow as the direction of propagation. Thus, a bias drain current could be necessary in certain SPW propagation applications. If the material quality is such that in the absence of drain current the plasmons can propagate tens of nm, then the impact of bias current on power dissipation would be negligible, since in principle *any nonzero* value of current would be sufficient to extend the propagation length. This aspect deserves further future study.

A prototypical NFL device, with the geometry of Fig. 1(b), could have the following parameters (see Appendix C): 1) 1.2 nm gate oxide; 2) 1200 Å n^+ poly gate; 3) n-type (MOSFET-like) sources and drains; 4) $W1 = W2 = 70.7$ nm; 5) $W3 = W4 = 30$ nm; 5) $W0 = 100$ nm; 6) $L0 = 200$ nm; 7) $L1 = 30$ nm; 7) $L2 = 50$ nm; 8) $\theta/2 \sim 30°$; 9) $\beta \sim 45°$. This prototypical NFL device would exhibit a threshold gate voltage to induce the 2DEG EF $U_0 \sim 0.35$ V [14]. Based on theoretical calculations for resonant plasma wave detection carried out by Knap *et al*. [14] for the photoresponse of a MOSFET with similar construction and gate length to those described earlier, at a gate voltage of approximately 1 V, the NFL would be capable of a switching frequency of approximately 6 THz at room temperature [14].

D. Plasmon–Plasmon Scattering

The high-speed and low-power consumption surrounding SPW launching and propagation, may be further exploited if, in addition to effecting guiding by tailoring the geometry of the patterned 2DEG, we cause the SPWs to interact so their direction of propagation is controlled by *active* means, in particular, by scattering one plasmon with another.

The key result, exploited by the NFL concept being presented, is the fact that the plasmon–plasmon interaction [17] is *repulsive* [18]. This fact is sufficient to posit the possibility that by controlling the generation and propagation (waveguiding) of individual plasmons, we can control their interaction, in particular, change the direction of a first plasmon by having a second one collide and transfer its momentum to it, in a desired spatial region of interaction. This is what was described pictorially in Fig. 2(b). For the sake of completeness, a derivation of the repulsive plasmon–plasmon interaction is made available in Appendix B.

IV. Circuits and Systems Considerations

An examination of the principles of operation of NFL, in light of circuits and systems considerations, see Fig. 3, elicits a number of questions. In particular:

1) What would be the ultimate device density?
2) What will be the ultimate device speed?
3) Is integration of many devices possible or limited by cross talk?
4) How to effect interconnections between gates?
5) How to interface NFL to conventional electronics?
6) How to "clock" these devices. We address these next.

A. Ultimate NFL Device Density

Once excited, SPWs will propagate until they reach the end of the waveguide/channel, where they are detected. In this operating regime, one would like to avoid SPW reflection within the channel (source–drain) cavity. Therefore, the device density will be limited by that at which resonance would occur. As the resonance frequency is $\omega_0 = \pi s_{\rm SPW}/2L$, one can see that a tradeoff exists between operating frequency and channel length and device density. The ultimate device density would be that of the smallest possible plasmon, which is an electric dipole [8]. The smallest possible dipole is an atom [8]. Therefore, the ultimate density would be equal to the areal atomic density of the material utilized.

In light of current highly scaled CMOS technologies, however, NFL logic compares quite favorably when one realizes that, with the equivalent of just two FET transistors, see Fig. 1(b), it realizes the FF function, which requires eight FET transistors in CMOS [19]. Thus, in principle, given NFL's dimensional compatibility with highly scaled CMOS, it has the potential to implement the same function in one-quarter the area as CMOS, which suggests a greater *equivalent* device density than achievable with CMOS.

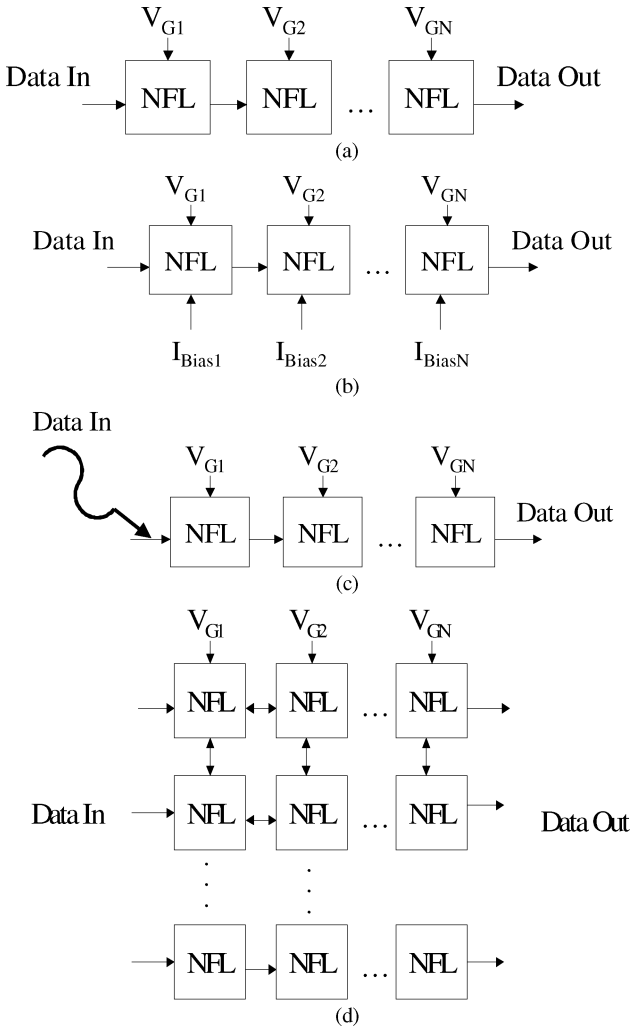

Fig. 3. (a) NFL circuit in which SPW propagation is gated by gate–source voltages V_{Gi} of subsequent gates. (b) NFL circuit in which SPW propagation is enhanced by bias currents. (c) NFL circuit in which SPWs are launched by light-plasmon coupling. (d) NFL array in which all of the earlier mentioned may be used for parallel data processing.

B. Ultimate NFL Device Speed

To assess what would control the ultimate NFL speed, one must reexamine the sequence of events leading to launching the SPWs, see Fig. 1(a). In particular, one can envision two modes of operation for the NFL device, namely, mode 1, in which the 2DEG EF is set up as an equilibrium state well in advance to the launching of the bias and control SPWs, and mode 2, in which the 2DEG EF is set up dynamically immediately prior to the launching of the bias and control SPWs. These modes are examined in detailed next.

1) Ultimate NFL Device Speed—Mode 1: In this mode, *prior* to launching the "bias" and "control" SPWs, the 2DEG EF would be created under the patterned gate by the application of the proper gate voltage; this would result in a *bias* state, namely, a charge sheet equilibrium (neutral) standby state that can *support* the propagation of SPWs, *prior* to the beginning of any logic operation. This state may be set at, e.g., the power up of the system, and the speed in question of any subsequent logic operation would be determined by the speeds at which the SPWs may be generated and propagated. These speeds are dictated by how fast can the equilibrium electron density be displaced, relative to the positive background charge, and the SPW propagation velocity.

The speed at which the equilibrium 2DEG EF density can be displaced equals the time it takes it to sense the appearance of, say, a negative charge on the source end, created by the departure from neutrality, on this end, due to the applied source–drain bias. This momentary departure from neutrality signifies that the 2DEG EF has begun to displace relative to the background positive charge, i.e., an SPW has been launched. Thus, the speed of SPW generation is that of SPW displacement in one direction, which is $\tau_{\mathrm{SPW\_Displacement}} = \pi/\omega_0$. If the source–drain distance is L, then the smallest switching time will be approximately given $\tau_{\mathrm{SPW\_Displacement}} + L/S_{\mathrm{SPW}}$.

The presence of a bias source–drain potential difference, which could be set up prior to any logic operation, if desired, establishes a baseline drift velocity v_0 that may extend the life/propagation distance of the "bias" and "control" SPWs. This current may be set up well in advance of effecting any logic operation; therefore, it does not slow down the device speed.

2) Ultimate NFL Device Speed—Mode 2: In this mode of operation, the 2DEG EF is set up immediately prior to launching the bias and control SPWs. The device speed, then, would be controlled by the parasitic effects and time constants associated with charging the NFL gate capacitance. This mode of operation, and the potential speed penalty incurred, however, would be justified only whenever it became necessary to interface with CMOS logic. The preferred mode of operation of NFL logic is self-timed, see shortly in Section IV-F.

C. Is Integration of Many Devices Possible or Limited by Crosstalk?

Integration of many devices is indeed possible. Since the 2DEG EF is created under highly conductive gates, these shield their vertical electric field, making gate-to-gate and 2DEG EF-to-2DEG EF coupling negligible. Also, since the plasmon motion is longitudinal, so is the electric field accompanying it. Therefore, lateral coupling is unlikely to be strong. This allows a high device density, only limited by fabrication techniques, not crosstalk. Further work, however, is needed to quantify these observations.

On the other hand, there will be coupling of devices connected in tandem. As discussed previously, in Section III-C, once generated, an SPW can propagate even in the absence of a source–drain potential difference. This means that an incoming SPW from a first NFL could also trigger the charge sheet displacement in a second NFL to which it is connected, i.e., the launching of another SPW. But this coupling is desirable as it is a mechanism for one NFL gate to drive another.

D. How to Effect Interconnections Between Gates?

The presence of SPWs may be detected in a number of ways [20]. These include the electric field produced by the oscillating plasmons, which may excite plasmons in adjacent

devices connected in tandem [21], and the transport of SPWs by a biasing current passing through multiple devices. Also, devices designed to operate near resonance may couple their excess radiation escaping the channel cavity to other devices.

E. How to Interface NFL to Conventional Electronics?

The interfacing of NFL to conventional electronics is straightforward. Since the gate–source bias is a voltage, this is a natural interface. Other possibilities include optoelectronic interfaces, where electron currents might be converted to light, and this light might in turn be coupled to the plasmons [22], and vice versa.

F. How to "Clock" NFL Circuits?

A little reflection on its principles of operation would reveal that NFL is a distributed device; it may be visualized as a medium through which waves, namely, plasmons, propagate. In this context, it lends itself very closely to *asynchronous logic design styles* [23]. Asynchronous digital design is a well-established digital design methodology, which has been discussed extensively by van Berkel *et al*. [23]. A number of useful properties germane to it include: high performance, low power, improved noise and electromagnetic compatibility properties, and a natural match with heterogeneous system timing. Asynchronous logic is self-timed; it requires no "clock." If one insists in clocking NFL circuits, then due to its femtosecond speeds, highest throughput is likely to require clocking via optoelectronic techniques.

V. Conclusion

We have presented the principles of NFL, a new digital "electronics" concept based on the controlled/active manipulation of SPs. NFL gates have the potential for *sub-fJ* power dissipation levels and *fs* switching speeds while operating at finite temperatures.

Appendix A

In this appendix, we define SPWs [20]. We begin with the concept of plasmon, which emerges from a consideration of the motion of a concentration $n(\vec{r},t)$ of free electrons, in a positive background n_0, as a result of an applied electric field $\vec{E}$. In particular, assuming that the electrons behave as a fluid of velocity $v(\vec{r},t)$, their motion is prescribed by the consistent solution of Newton's and the continuity equations,

$$m\frac{d\vec{v}}{dt} + m(\vec{v}\cdot\nabla)\vec{v} = -e\vec{E} \tag{A1}$$

and

$$\frac{\partial n}{\partial t} + \nabla\cdot(n\vec{v}) = 0. \tag{A2}$$

As a first step toward the solution, after neglecting the second term in (A1) due to its quadratic nature in $\vec{v}$, one postulates that the effect of the electric field is to cause the local electron density to deviate from the constant background density by $\delta n = n - n_0$. In this context, the extent of this deviation is related to the electric field by Poisson's equation,

$$\nabla\cdot\vec{E} = -4\pi e(n-n_0) = -4\pi e\delta n \tag{A3}$$

and, because of electron inertia and the restoring force supplied by Coulomb attraction to regain equilibrium, i.e., $\delta n = 0$, oscillations ensue. These collective bulk electron oscillations are called volume plasmons, and their frequency of oscillation is obtained by substitution of δn into (A2), resulting in,

$$\frac{\partial\delta n}{\partial t} + n_0\nabla\cdot\vec{v} = 0 \tag{A4}$$

which, upon differentiating with respect to time, becomes,

$$\frac{\partial^2\delta n}{\partial t^2} + n_0\nabla\cdot\frac{\partial}{\partial t}\vec{v} = \frac{\partial^2\delta n}{\partial t^2} + n_0\nabla\cdot\left(\frac{-e\vec{E}}{m}\right) = 0 \tag{A5}$$

and which, in turn, upon substituting (A3) into (A5) becomes,

$$\frac{\partial^2\delta n}{\partial t^2} + \frac{4\pi e^2 n_0\delta n}{m} = 0. \tag{A6}$$

(A6), being analogous to that of a harmonic oscillator, prescribes the frequency of plasmon oscillation as

$$\omega_p = \sqrt{\frac{4\pi n_0 e^2}{m}}. \tag{A7}$$

SPs, Fig. 4, thoroughly reviewed by Raether [22], are elicited by the interaction of external electromagnetic surface waves with *surface* electrons and are characterized by a dispersion relation, a spatial extension, and a propagation length or lifetime [3], [4], [20], [22]. In particular, when Laplace equation $\nabla^2\phi = 0$ is solved at the *interface* between a region of negative dielectric constant, such as a metal, and a dielectric, Fig. 4, one obtains solutions (surface waves) that propagate along the interface and decay exponentially away from it [20], [22]. Such propagating waves, which couple a propagating surface electromagnetic wave with surface carriers are called SPWs [3], [4]. In our device, the metal–dielectric interface is realized by, e.g., the electron fluid under an MOS gate, and the gate oxide.

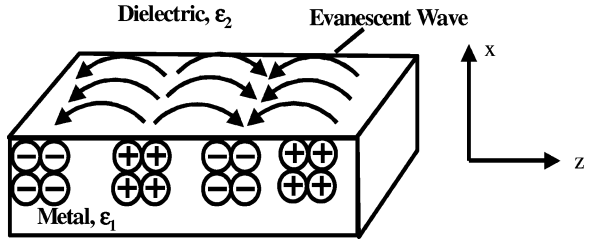

Fig. 4. Sketch of surface plasmon. The field accompanying a surface plasmon peaks at the dielectric–metal interface and diminishes exponentially away from the interface. *After* [20].

Appendix B

In this appendix, we summarize a result exploited in our proposed device concept, namely, that plasmon–plasmon interaction is repulsive [18] and that by controlling the generation and propagation (waveguiding) of individual plasmons, we can control their interaction, in particular, change the direction of

a first plasmon by having a second one collide and transfer its momentum to it, in a desired spatial region of interaction. We also provide a proof for plasmon–plasmon repulsive interaction at finite temperatures.

The plasmon excitations in an electron gas are in general described by the Hamiltonian [17], (B1).

$$H = \sum_p E_p c_p^+ c_p + {\sum_k}' \omega_0 \left(b_k + \frac{1}{2} \right) + H_{\rm III} + H_{\rm IV} + H_{\rm SR} \tag{B1}$$

where $p = (\vec{p}, \sigma)$, $E_p = p^2/2m^*$ is the free-electron energy, σ is the electron spin, ω_0 is the plasma frequency, and units of $\hbar = 1$ have been assumed. The prime on the summation sign is meant to convey that $k < k_c$, where k_c is a cutoff wave vector below which the plasmons are well defined, and c_p^+ and c_p are the electron creation and annihilation operators, which obey the usual anticommutation relations, and the operators b_k^+ and b_k correspond, respectively, to the creation and annihilation of a plasmon excitation and also obey the usual commutation rules [17].

The interaction terms are given below by (B2)–(B4).

$$H_{\rm III} = -i \sum_k \sum_p \gamma_3 \left(k, p \right) \pi_k c_p^+ c_{p+k} \tag{B2}$$

where $\gamma_3(k,p) = (2\pi e^2/\omega_0 k^2)^{1/2} (E_{k+p} - E_k)$ describes the linear coupling of electrons and plasmons, and $\pi_k = b_{-k} + b_k^+$.

$$H_{\rm IV} = -{\sum_k}' {\sum_{\substack{k \\ (k+k' \neq 0)}}}' \sum_p \gamma_4 \left(k, k' \right) \pi_k \pi_{k'} c_p^+ c_{p+k+k'} \tag{B3}$$

describes the bilinear coupling between electrons and plasmons, where $\gamma_4 \left(k, k' \right) = \pi e^2 \cos\theta_{kk'}/m^* \omega_0$, with $\theta_{kk'}$ representing the angle between $\vec{k}$ and k'.

$$H_{\rm SR} = \frac{1}{2} \sum_{k > k_c} \sum_p \sum_{p'} \frac{4\pi e^2}{k^2} c_{p+k}^+ c_{p'-k}^+ c_{p'} c_p \tag{B4}$$

represents a residual short-range interaction between electrons of importance in the definition of the electron propagator.

The evolution of two plasmons, leading to their interaction, is captured by the two-plasmon Green's function, which is defined by [17],

$$\begin{aligned} &D_2(k_1 t_1 k_2 t_2; k_1' t_1' k_2' t_2') \\ &\quad = \left[(-i)^2/4\omega_0^2 \right] \left\langle T\left(\pi_{k_1}(t_1) \pi_{k_2}(t_2) \pi_{k_2'}^+(t_2') \pi_{k_1'}^+(t_1') \right) \right\rangle \end{aligned} \tag{B5}$$

where T is the Dyson time-ordering symbol [24], [25] with $k_1 + k_2 = k_1' + k_2' \neq 0$. This, in turn, is related to the one-plasmon Green's function defined by,

$$D_1(k_1 t_1; k_1' t_1') = [(-i)/2\omega_0] \left\langle T\left(\pi_{k_1}(t_1) \pi_{k_1'}^+(t_1') \right) \right\rangle. \tag{B6}$$

and the one-electron Green's function G is defined by

$$G(pt; p't') = (-i) \left\langle T\left(c_p(t) c_{p'}^+(t') \right) \right\rangle. \tag{B7}$$

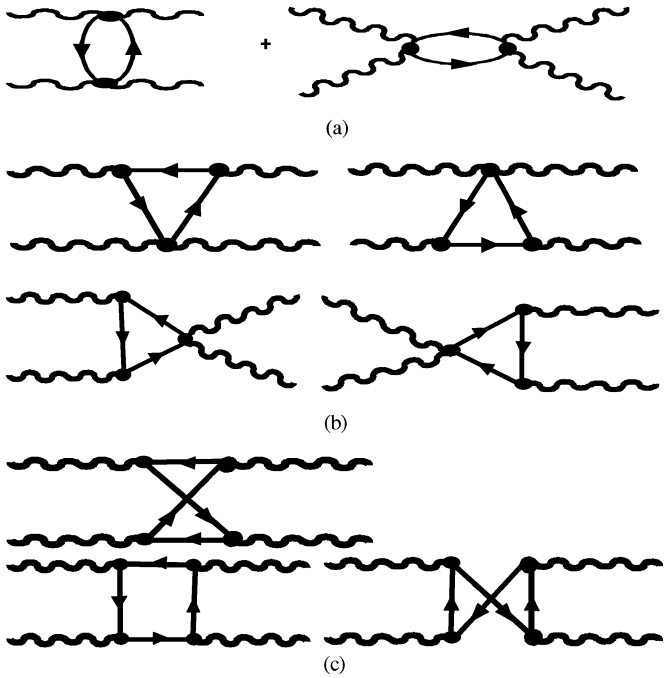

Fig. 5. Diagrammatic representation of the scattering of two plasmons (wavy lines). Solid lines denote electrons (→) and holes (←), respectively. Dots denote the Bohm–Pines vertices for electron–plasmon interactions [17]. (a) Bubble Contributions. (b) Triangle Contributions. (c) Square Contributions.

Now, in terms of the evolution of the one-plasmon propagators, the two-plasmon propagator is given by [17]

$$\begin{aligned} D_2(k_1 k_2; k_1' k_2') &= D_2^{(0)}(k_1 k_2; k_1' k_2') \\ &\quad + \iint D_1^{(0)}(k_1) D_1^{(0)}(k_2) 4\omega_0^2 K(k_1 k_2; k_3 k_4) \\ &\quad \times D_2(k_3 k_4; k_1' k_2') \frac{d^4 k_3}{(2\pi)^4} \frac{d^4 k_4}{(2\pi)^4} \end{aligned} \tag{B8}$$

where $k_1 \equiv (\vec{k}_1 k_1^0)$, $D_2^{(0)}$ is the noninteracting two-plasmon propagator, given by, $D_1^{(0)}(k_1) D_1^{(0)}(k_2)(\delta_{k_1 k_1'} \delta_{k_2 k_2'} + \delta_{k_1 k_2'} \delta_{k_2 k_1'})$, and $D_1^{(0)}(k_1)$ is the one-plasmon propagator:

$$D_1^{(0)}(k_1) = \frac{1}{\lfloor (k_1^0)^2 - \omega_0^2 \rfloor}. \tag{B9}$$

The importance of expressing the two-plasmon propagator in terms of the one-plasmon propagator is enormous; it underpins the idea of controlling the interaction between two plasmons. In particular, (B8) prescribes that these plasmons may be produced independently and then guided to the spatial region where they will be made to interact. This is the essence of our device concept, as shown in Fig. 2, where the independent plasmons propagating along $\vec{k}_{\rm Bias}$ and $\vec{k}_{C2}$ are made to proceed toward the junction, where they interact.

The interaction between plasmons is captured by the kernel $K\left(k_1 k_2; k_3 k_4\right)$, which is visualized by the diagrammatic representation, as shown in Fig. 5 [17].

Accordingly, it is expressed as

$$K = K_B + K_T + K_S \tag{B10}$$

where K_B stands for terms arising from two γ_4's (B for bubbles), K_T stands for terms arising from one γ_4, and two γ_3's (T for triangles), and K_S stands for terms arising from four γ_3's (S for square). Out of these contributions to the kernel, it turns out that the triangle and square parts are negligible with respect to the bubble contributions [17]. The essence of the bubble contributions is given by the strength of the plasmon–plasmon coupling,

$$g_4(Q) \cong 4\omega_0^2 K_B(Q+k,-k;Q+k',-k') \tag{B11}$$

where Q is the total momentum [17]. An expression for $g_4(Q,\omega)$ has been obtained by Mandal and Tripathy [18] as,

$$\begin{aligned} g_4(Q,\omega) &= \frac{1}{8[F_0(Q,\omega)]^2} \\ &\times \sum_{\substack{p_1<k_c \\ |p_1+Q|<k_c}} \frac{128}{\omega^2-[\omega(p_1)+\omega(p_1+Q)]^2} \\ &\times \sum_{\substack{k'<k_c \\ |Q-k'|<k_c \\ (p_1)+k'\neq 0}} \gamma_4(p_1,k')\gamma_4(-p_1-Q,Q-k') \\ &\times \{[\omega-\omega(k')-\omega(Q-k')]^{-1}-[\omega+\omega(k')+\omega(Q-k')]^{-1}\} \\ &\times \sum_{p\sigma} n_{p,\sigma}(1-n_{p+p_1+k',o}) \end{aligned} \tag{B12}$$

where $n_{p,\sigma}$ are the occupation numbers appropriate to the zero-temperature plasma [26] and the plasmon dispersion relationship is given by (B13) [18].

$$\omega(k) = \omega_0 + \sum_{p\sigma} [\gamma_3(k,p)]^2 \frac{n_{p,\sigma}-n_{p+k,\sigma}}{\omega_0 - E_{p+k}+E_p} \tag{B13}$$

Employing the random phase approximation (RPA) [25], [26], evaluation of (B18) yields [18]

$$\omega(k) = \omega_0 + k^2/2\mu, \qquad \mu = 5m^*\omega_0/6E_F$$

where E_F is the Fermi energy. Using this dispersion relation, evaluation of (B13) for the case $Q=0$ and $\omega = 2\omega_0$, yields [18]

$$\begin{aligned} g_4(Q=0,\omega=2\omega_0) &= \frac{\mu k_F}{F_0^2}\frac{5\alpha r_s}{192\pi^5} \times \int_{\substack{k'<\beta \\ |p'-k'|<\beta}} dp'dk' \frac{[(p'-k')\cdot k']^2}{|p'-k'|^4\, k'^2} \\ &\times \left(\frac{1}{k'^2}+\frac{1}{k'^2+5\omega_0^2/3}\right) S_{\mathrm{HF}}(p') \end{aligned} \tag{B14}$$

where

$$\begin{aligned} F_0(Q=0,\omega=2\omega_0) &= \left(\mu k_F/\pi^2\right) \\ &\times \left[-\beta+\sqrt{5\omega_0^2/12}\tan^{-1}\left(\beta/\sqrt{5\omega_0^2/3}\right)\right] \end{aligned} \tag{B15}$$

and $\beta = k_c/k_F$. $S_{\mathrm{HF}}(p')$ denotes the Hartree–Fock structure factor [27], and $r_s = (3/4\pi n_s)^{1/3}/a_0$. $a_0 = \hbar^2/me^2$, and $\alpha \cong 1/137$ is the fine structure constant. Using (B14), Mandal and Tripathy [18] calculated g_4 for various values of r_s from 1 to 10, Fig. 6, and determined that *the sign of g_4 is positive for all r_s*, thereby proving that *the plasmon–plasmon interaction is repulsive!* SPWs interacting in the "active" region/junction of our device will be steered/deflected as prescribed by momentum conservation principles.

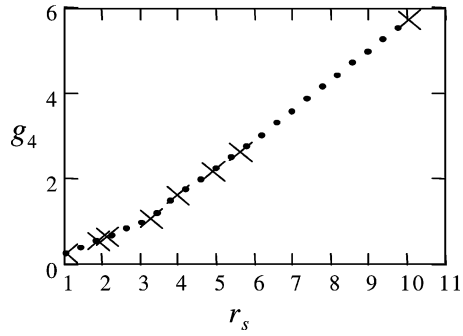

Fig. 6. Plasmon–plasmon scattering strength, in units of $1/\mu k_F$, as function of normalized volume per conduction electron. *After* [18].

Since the earlier analysis assumed zero absolute temperature, the final result, namely, that the plasmon–plasmon scattering strength remains *repulsive,* even at finite temperatures, may elicit some skepticism. A number of arguments to support the belief that g_4 remains repulsive are as follows.

First, as pointed out by Pines [26], the *operator* approximation nature of the RPA allows the effect of the temperature to enter the analysis at the end of the calculation by replacing the *operators* $c_{p\sigma}^+ c_{p\sigma}$ and $c_{p+k\sigma}^+ c_{p+k\sigma}$ with their expectation values $n_{p,\sigma}$ and $n_{p+k,\sigma}$ at zero temperature, and these, in turn, by the corresponding occupation numbers $f_{p,\sigma}$ appropriate at finite temperatures [26], namely,

$$f_{p,\sigma} = \frac{1}{e^{(\varepsilon(p)-E_F)/k_BT}+1} \tag{B16}$$

where $\varepsilon(p)$ is the single-particle energy, and k_B is Boltzmann's constant. Insertion of (B16) into (B13) yields the finite-temperature plasmon dispersion,

$$\omega(k) = \omega_0 + \sum_{p\sigma} [\gamma_3(k,p)]^2 \frac{f_{p,\sigma}-f_{p+k,\sigma}}{\omega_0 - E_{p+k}+E_p} \tag{B17}$$

but under the RPA this may again be approximated by $\omega(k) = \omega_0 + k^2/2\mu$ [18]. Therefore, no changes in the *repulsive* nature of the plasmon–plasmon interaction should be caused by operation at finite temperatures.

Second, a mathematical proof that deducts g_4 remains repulsive at finite temperatures may be given as follows:

Lemma 1: At zero temperature, the plasmon–plasmon interaction strength g_4 is repulsive (see Fig. 6).

Lemma 2: At zero temperature, the plasmon–plasmon interaction strength is given by (B18).

$$g_4(Q) \cong 4\omega_0^2 K_B(Q+k,-k;Q+k',-k') \tag{B18}$$

Lemma 3: The plasmon–plasmon interaction strength g_4 will be repulsive if $K_B > 0$.

Lemma 4: The bubble contribution K_B is given by (B19).

$$\begin{aligned} K_B &= 4\gamma_4(k_1,k_2)\gamma_4(k_3,k_4)\pi_0(k_1+k_2) \\ &+ 4\gamma_4(k_1,k_3)\gamma_4(k_2,k_4)\pi_0(k_1-k_3) \\ &+ 4\gamma_4(k_1,k_4)\gamma_4(k_2,k_3)\pi_0(k_2-k_3) \end{aligned} \tag{B19}$$

where

$$\pi_0(k) = \int \frac{d^3p}{(2\pi)^3} G_0(p+k)G_0(p) \tag{B20}$$

with G_0 the zero-temperature single-particle free propagator Green's function [24].

Lemma 5: K_B is made up of three terms, each one of the same form, namely, $4\gamma_4(k_1,k_2)\gamma_4(k_3,k_4)\pi_0(k_1+k_2)$.

Lemma 6: If each of the terms in K_B is greater than zero, then g_4 will be nonnegative and the plasmon–plasmon interaction repulsive.

Lemma 7: If the angle between the momenta of the interacting plasmons is $0 < \theta_{kk'} \leq \pi/2$, then the factor, $\gamma_4(k,k') = \pi e^2 \cos\theta_{kk'}/m^*\omega_0$ is positive at any temperature.

Lemma 8: The function $f^-(\omega) = [1+e^{\omega/k_BT}]^{-1} > 0$ for any and all values of ω and T.

Lemma 9: The function $f^+(\omega) = 1-[1+e^{\omega/k_BT}]^{-1} > 0$ for any and all values of all ω and T.

Lemma 10: Due to their exponential dependence, the dominant temperature dependence of g_4 is limited to that of the Fermi functions *f*.

Lemma 11: The integral of a function must be nonnegative if the function is nonnegative.

Theorem: The plasmon–plasmon interaction strength will be repulsive at any temperature, if $g_4 > 0$ at any temperature.

Proof:

Step 1: By Lemmas 1–7, at any temperature each of the terms in K_B will be positive if the finite-temperature version of the factor π_0, namely, π_0^T, is positive at any temperature.

Step 2: π_0^T, according to Mattuck [24], is given by,

$$\pi_0^T(k,\omega) = f^+(\omega)\Pi_0(k,\omega+i\delta) + f^-(\omega)\Pi_0(k,\omega-i\delta) \tag{B21}$$

where

$$\Pi_0(k,\omega_i) = 2\int \frac{d^3p}{(2\pi)^3} f_p(1-f_{p+k})\left[\frac{1}{i\omega_i - E_p + E_{p+k}} - \frac{1}{i\omega_i - E_{p+k} + E_p}\right]. \tag{B22}$$

and, for applications at *real* frequencies, the ones of interest to us, $i\omega_i$ is analytically continued to $\omega+i\delta$ and $\omega-i\delta$ [24]. In (B22), f_p is given by

$$f_p \equiv f(E_p - E_F) = \frac{1}{1+e^{\frac{E_p-E_F}{k_BT}}} \tag{B23}$$

where $E_p = \hbar^2|\vec{p}|^2/2m^*$, E_F is the Fermi energy. Explicitly, π_0^T may be expressed as in (B24) [24].

$$\pi_0^T(k,\omega) = \int_{\substack{E_p<E_F \\ E_{p+k}>E_F}} \frac{d^3p}{(2\pi)^3} \left\{ \begin{aligned} &f^+(\omega)\left[\frac{1}{\omega - E_p + E_{p+k} - i\delta} - \frac{1}{\omega - E_{p+k} + E_p + i\delta}\right] \\ &+f^-(\omega)\left[\frac{1}{\omega - E_p + E_{p+k} - i\delta} - \frac{1}{\omega - E_{p+k} + E_p + i\delta}\right] \end{aligned} \right\} \tag{B24}$$

Step 3: By Lemmas 8 and 9, the factors $f^-(\omega)$, $f^+(\omega)$, $f_p(1-f_{p+k})$ are nonnegative at any temperature.

Step 4: For a fixed Fermi level, taking the $T \to 0$ limit on π_0^T yields (B25) [24].

$$\pi_0^{T\to 0}(k,\omega) = \int_{\substack{E_p<E_F \\ E_{p+k}>E_F}} \frac{d^3p}{(2\pi)^3} \left\{ \frac{1}{\omega - E_p + E_{p+k} - i\delta} - \frac{1}{\omega - E_{p+k} + E_p + i\delta} \right\} \tag{B25}$$

Then, since, by Lemma 10, (B25) contains no Fermi functions f, it is temperature-independent. Therefore, at zero temperature, g_4 may also be expressed as

$$g_4(Q) \cong 4\omega_0^2 K_B^{T\to 0}(Q+k,-k;Q+k',-k') \tag{B26}$$

where

$$\begin{aligned} K_B^{T\to 0} &= 4\gamma_4(k_1,k_2)\gamma_4(k_3,k_4)\pi_0^{T\to 0}(k_1+k_2) \\ &+ 4\gamma_4(k_1,k_3)\gamma_4(k_2,k_4)\pi_0^{T\to 0}(k_1-k_3) \\ &+ 4\gamma_4(k_1,k_4)\gamma_4(k_2,k_3)\pi_0^{T\to 0}(k_2-k_3) \end{aligned} \tag{B27}$$

Step 5: At zero temperature, and the same values of momentum and frequency, (B26) must be equal to (B14). But, since the *integrand* of (B14) is nonnegative, by Lemma 11, the integrand embodied in (B26), namely, that in (B25), must also be nonnegative.

Step 6: Furthermore, the integrand in (B25), (in $\pi_0^{T\to 0}$ at zero temperature) is the same as those coupled to $f^-(\omega)$ and $f^+(\omega)$ in the integrand of π_0^T at finite temperature (B24). Therefore, these integrands, by Step 3 are not only nonnegative but also remain so even at finite temperatures, since the temperature dependence is in the Fermi function factors, not in them (the integrands), and by Step 5, we know these Fermi functions are nonnegative at any temperature.

Step 7: Therefore, since at zero temperature g_4 is nonnegative, by Step 6 it must also be nonzero at any finite temperature. The theorem is thus proved.

The restriction that $0 < \theta_{kk'} \leq \pi/2$ is not believed to be restrictive and, therefore, we deem it reasonable to expect that the theorem will hold in the most cases of practical interest.

APPENDIX C

In this appendix, we address the physical factors underlying NFL design and geometry.

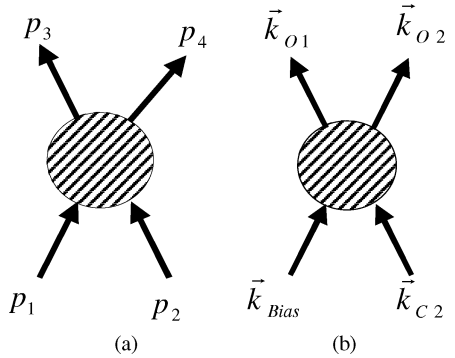

Fig. 7 (a) Kinematics of a general scattering process. (b) Kinematics of SPW scattering in NFL.

A. Background

Conceptually, NFL derives from, and embodies, the prototypical scattering process, Fig. 7(a) [28].

In the general scattering process, e.g., two particles, with momentum p_1 and p_2, are scattered off a target. Then, depending on the nature of the target, the collision process may involve mere scattering, or it can result in absorption or creation of new particles, with the overall process obeying fundamental conservation laws of energy, momentum, angular momentum, parity, charge conjugation, internal symmetries, etc [28].

B. Physics Underlying Scattering in NFL

The NFL embodies and exploits the fundamental question in a scattering process, which may be posited as follows: given an initial state consisting of two incoming particles evolving freely and isolated from each other, moving with momentum p_1 and p_2, respectively, what would be their outgoing momentum states, say, p_3 and p_4 (assuming no particle absorption or creation), long after they interact and they are again described by free particle kinematics? The answer to this question is derived in quantum field theory (QFT) [28] and is related to the quantum mechanical transition amplitude $\langle f, \text{out} \mid i, \text{in}\rangle$, which enables obtaining the probability that the incoming state $\mid i, \text{in}\rangle$will evolve in time and be measured in the outgoing state $\mid i, \text{out}\rangle$, namely, the transition probability $W_{f\leftarrow i} = |\langle f, \text{out} \mid i, \text{in}\rangle|^2$.

The pertinent quantity, which expresses the relation between the transition amplitude and actual measurements, however, is the *differential scattering cross section* $d\sigma$ for scattering to occur in the solid angle (θ, ϕ). $d\sigma$ is defined as the transition probability per scatterer in the target and per unit incident flux [28], [29]. As is well known, QFT provides a recipe for computing $d\sigma$ based on the following lemmas [28].

Lemma 1: In the absence of external sources, space becomes homogeneous and translation invariance exists.

Lemma 2: In a translation invariant system the matrix elements of $W_{f\leftarrow i}$ vanish unless energy and momentum are conserved.

Lemma 3: The total probability of transitioning from the input state to all states is unity.

Lemma 4: There exists an isomorphism among the in- and out-Fock space and the out-state may be expressed in terms of the in-state.

Lemma 5: The isomorphism between the out- and in-states implies the existence of a unitary operator S, the so-called "S" matrix of quantum field theory, which is related to the transition amplitude by

$$\langle f, \text{out} \mid i, \text{in}\rangle = \langle f, \text{in}| \, S \, |i, \text{in}\rangle \tag{C1}$$

where [29]

$$\langle f| \, S - 1 \, |i\rangle = iA_{fi} \, (2\pi)^4 \, \delta^{(4)} \, (p_f - p_i) \, . \tag{C2}$$

In (C2), p_i and p_f are the sum of the initial and final four-momenta, respectively, and the factor i in front is a convention, which is there to match nonrelativistic quantum mechanics [29]. A_{fi} is the scattering amplitude obtained from the Feynman diagrams [24], [25] for the expansion $\langle f|S-1|i\rangle$.

Based on these assumptions, the differential scattering cross section in the center-of-mass reference frame is prescribed in QFT [28], [29] by the formula:

$$d\sigma(\theta,\phi) = \left[\frac{1}{4E_1E_2|\vec{v}_1 - \bar{v}_2|}(2\pi)^4\delta^{(4)}(p_f - p_i) \prod_{\text{final states}} \times \frac{d^3p_i}{(2\pi)^3}\frac{1}{2E_{\bar{p}_i}}\right] \cdot |A_{fi}|^2 \tag{C3}$$

where E_1 and E_2 are the energies of the incoming particles and v_1 and v_2 are their respective velocities, and A_{fi} captures the interactions.

The following observations may be drawn from the aforementioned equation (C3). First, because of the delta function, both momentum and energy are conserved during scattering [30]. Second, $d\sigma$ is determined by two factors, namely, a kinematic one and a quantum-field-theoretical one. The former is embodied by the first factor in (C3), whereas the latter is embodied by $|A_{fi}|^2$.

It can be shown [28], [29] that evaluation of $d\sigma$ in the center-of-mass reference frame yields a constant times $|A_{fi}|^2$. Therefore, to determine the *angular* dependence of $d\sigma$ it suffices to evaluate $|A_{fi}|^2$.

For the general scattering event involving two incoming plasmons with momentum vectors k_1 and k_2, interacting with each other in a high-density electron gas and being scattered into two outgoing plasmons with momentum vectors k_3 and k_4, the scattering amplitude was obtained by Rajagopal *et al.* [17] as

$$\begin{aligned} A_{fi} = {} & 4\gamma_4(k_1,k_2)\gamma_4(k_3,k_4)\pi_0(k_1+k_2) \\ & + 4\gamma_4(k_1,k_3)\gamma_4(k_2,k_4)\pi_0(k_1-k_3) \\ & + 4\gamma_4(k_1,k_4)\gamma_4(k_2,k_3)\pi_0(k_2-k_3) \end{aligned} \tag{C4}$$

where

$$\pi_0(k) = \int \frac{d^3p}{(2\pi)^3} G_0(p+k)G_0(p) \tag{C5}$$

with G_0 is the single-particle free-propagator Green's function. Designing the NFL, in particular, choosing the angles between

the channels supporting the propagating *bias* and *control* plasmons, and the scattered plasmons requires knowledge of the angular variation of the $|A_{fi}|^2$ factor of $d\sigma$. We calculate this next.

C. Angular Dependence of Differential Scattering Cross Section

In this section, we address how to determine the angular dimensions in the NFL. The surface plasmons propagating in the NFL do so while confined to *two* dimensions. In this context, the total scattering cross section is defined by a "length," not an area, and the angular variation is given not by a solid angle, but by a single angle. But we are not interested in calculating the total scattering cross section; just the angular variation of the *differential* scattering cross section.

Now, the NFL mode of operation under consideration establishes the scattering situation depicted in Fig. 7(b), where $\vec{k}_1 = \vec{k}_{\rm Bias}$, $\vec{k}_2 = \vec{k}_{C2}$, $\vec{k}_3 = \vec{k}_{O1}$, and $\vec{k}_4 = \vec{k}_{O2}$. Therefore, by design, i.e., by the topology of our device, we only employ the first term in (C4) and this determines A_{fi}. Explicitly, then, with $\gamma_4\,(k,k') = \pi e^2 \cos\theta_{kk'}/m^*\omega_0$ [17] we can express A_{fi} as

$$A_{fi} = 4\left(\frac{\pi^2 e^2}{m^*\omega_0}\right)^2 \cos\theta_{k_1k_2}\cos\theta_{k_3k_4}\,\pi_0\,(k_1+k_2) \qquad \text{(C6)}$$

where the factor $\cos\theta_{k_1k_2}\,\pi_0\,(k_1+k_2)$ captures the differential cross section's dependence on the angle between the incoming particles (plasmons in our case), and $\cos\theta_{k_3k_4}$ its dependence on the angle of the outgoing particles, the position of the measuring "detector." In other words, we utilize the differential cross section to determine how to best choose the angles between $\vec{k}_{\rm Bias}$ and $\vec{k}_{C2}$, and $\vec{k}_{O1}$ and $\vec{k}_{O2}$.

To completely determine the angular dependence, we have to evaluate the factor $\pi_0(k_1+k_2)$. This has been done by Mattuck [24], who obtained the following expression for it at a *finite temperature*:

$$\pi_0^T\,(k,\omega) = f^+(\omega)\pi_0(k,\omega+i\delta) + f^-(\omega)\pi_0(k,\omega-i\delta) \qquad \text{(C7)}$$

where $f^-(\omega) = [1+e^{\omega/k_BT}]^{-1}$ and $f^+(\omega) = 1-f^-(\omega)$, and in two dimensions we have,

$$\Pi_0(k,\omega_i) = 2\int \frac{d^2p}{(2\pi)^2} f_p(1-f_{p+k})\left[\frac{1}{i\omega_i - E_p + E_{p+k}} - \frac{1}{i\omega_i - E_{p+k} + E_p}\right] \qquad \text{(C8)}$$

with

$$f_p \equiv f(E_p - E_F) = \frac{1}{1+e^{\frac{E_p-E_F}{k_BT}}} \qquad \text{(C9)}$$

where $E_p = \hbar^2|\vec{p}|^2/2m^*$, E_F is the Fermi energy and k_B is Boltzmann constant and T is absolute temperature.

Therefore, the angular dependence of the differential scattering cross section may be written as,

$$\frac{d\sigma\,(\theta)}{d\theta} = C\left[\cos(\theta_{k_1k_2})\cos(\theta_{k_3k_4})\Pi_0^T\,(k_1+k_2)\right]^2 \qquad \text{(C10)}$$

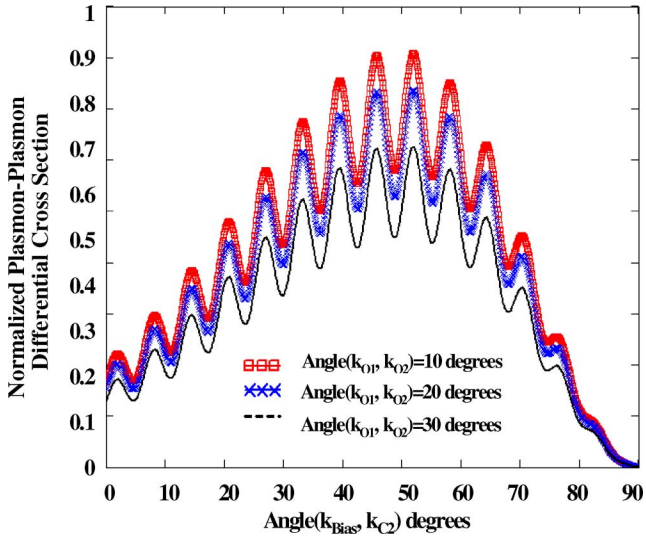

Fig. 8. Normalized differential scattering cross section for various incoming angles and set detector angles. Assumptions: In the coordinate system employed the momenta are defined as follows: $\vec{p} = |\vec{p}|\cdot\hat{j}$, $\vec{k}_1 = |\vec{k}_1|\cdot\hat{j}$, $\vec{k}_2 = -|\vec{k}_2|\cos(\theta_{k_1k_2})\cdot\hat{i} + |\vec{k}_2|\sin(\theta_{k_1k_2})\cdot\hat{j}$, $|\vec{k}_1| = |\vec{k}_2| = 0.5k_F$, $|k_F = \sqrt{2\pi n_s}|$, $m^* = 1.08m_0$, $\omega = \omega_0 = 7.022\times10^{13}\ \mathrm{s}^{-1}$, $n_s = 6.295\times10^{16}\ \mathrm{m}^{-2}$, $\varepsilon_r = 11.8$, $T = 300\ \mathrm{K}$. The integration limits over p in (C8) are $0.01k_F$ and k_F.

Fig. 8 shows the computed normalized plasmon–plasmon differential scattering cross section at room temperature as function of angle between incoming plasmons $\vec{k}_{\rm Bias}$ and $\vec{k}_{C2}$. As can be seen from Fig. 8, the differential scattering cross section is an undulating function of the angle between the incoming plasmons. The undulating behavior is found to be a function of the relative magnitude of the plasmon momenta with respect to the electron momentum, there being deeper modulation for greater plasmon momenta. Due to potential fabrication limitations, it should be desirable to have as large an angle as possible, compatible with a sufficiently large differential scattering cross section. An angle between 30° and 60° seems a good compromise (see prototypical design at the end of Section III-C).

Similar considerations apply for defining the angle of the detecting channel receiving the desired outgoing plasmon $\vec{k}_{O1}$. Obviously, since there is no momentum in the direction opposite $\vec{k}_{C2}$, the angle of the output plasmon is measured with respect to $\vec{k}_{\rm Bias}$. The plot in Fig. 8 shows that for a given angle between incoming plasmons, $d\sigma$ is greater near the direction of $\vec{k}_{\rm Bias}$. The desired detection angle should be chosen as small as possible, as permitted by limitations imposed by fabrication and potential coupling with the other output channel $\vec{k}_{O2}$. Based on Fig. 8, an angle of 30° may be chosen for a prototypical NFL device (see end of Section III-C).

D. NFL Channel Lengths

In this section, we address how to determine the lengths of the plasmon-transporting channels.

As determined by Dyakonov and Shur [31], the velocity and channel potential accompanying a plasmon wave propagating

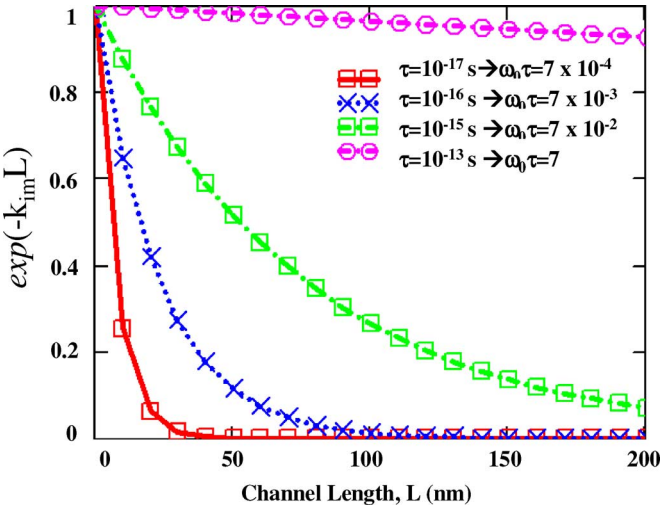

Fig. 9. Plasmon wave spatial profile along channel length for various plasmon damping times τ. Parameters: $\omega = \omega_0 = 7.022 \times 10^{13}\mathrm{s}^{-1}$, $\varepsilon_r = 11.8$, $d = 1.2$ nm, $U_0 = 0.35$ V, $C = \varepsilon_r\varepsilon_0/d$, $n_s = (CU_0/e) = 6.295 \times 10^{16}$ m^{-2}, $s = 1.372 \times 10^7$m/s, $7 \times 10^{-4} \leq \omega_0\tau \leq 7$.

in a direction "x" are proportional to,

$$\exp(ikx - \omega t) \tag{C11}$$

(we assume a very narrow plasmon wave packet centered at a given k) k is the propagation constant and possesses a dispersion relation given by [31]

$$k = \pm k_0 = \pm\frac{\omega}{s}\sqrt{1 + \frac{i}{\omega\tau}}. \tag{C12}$$

ω is the frequency and, in general, $1/\tau = 1/\tau_{ep} + 1/\tau_\nu$ is the inverse plasma damping time, where τ_{ep} is the electron collision time due to scattering with phonons and impurities and τ_ν, defined in Section III-C, is the viscosity of the electronic fluid. From an examination of (C12), we observe that when $\omega\tau \gg 1$, the propagation constant has a negligible imaginary part. Therefore, as per (C11), the plasmon wave would propagate with negligible attenuation. On the other hand, when $\omega\tau \ll 1$, the propagation constant has a dominant imaginary part. Therefore, as per (C11), it would be damped upon being launched. Since the $\omega\tau$ product is key, to determine the channel length we plot the spatial wave profile for as function on propagation length for various, $\omega\tau$ products. Fig. 9 indeed shows that for NFL operation, an $\omega\tau$ product of at least 7 would be required for detecting a plasmon at an amplitude of 20% its maximum 200 nm away from the input. This, in turn, could be achieved with a plasmon damping time of $\tau = 0.1\ ps$ and a plasmon frequency of $\omega_0 = 7.022 \times 10^{13}\mathrm{s}^{-1}$. For NFL applications, this product may be "tuned" by adjusting the gate bias voltage. Therefore, channel lengths under 200 nm may be chosen for an NFL device.

E. NFL Channel Widths and Junction

Plasma wave oscillations are longitudinal [8], and their direction of propagation is determined by the boundary conditions along the longitudinal direction, namely, those at the source and at the drain [3]. Once launched, the electron fluid induced under the patterned gate will guide the SPWs along the channel length. Given the wave nature of plasmons, the channel width is determined by various factors. These include, in particular, the need for minimizing impedance discontinuity-induced reflections [32] at the junction, the area where the *bias* and *control* SPWs interact. Thus, the channel width along its length may be variable (i.e., taper off from the initial width at the input). Other considerations are the desired source–drain current, and practical fabrication limitations (minimum realizable width). Based on current fabrication capabilities, the channel width could be as small as tens of nm [14].

ACKNOWLEDGMENT

The author would like to thank the anonymous reviewers for their encouragement and many useful suggestions.

REFERENCES

[1] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, no. 8, pp. 56–59, Apr. 19, 1965.
[2] Y. Taur, "CMOS design near the limit of scaling," *IBM J. Res. Dev.*, vol. 46, no. 2/3, pp. 3–7, 2002.
[3] M. I. Dyakonov and M. S. Shur, "Shallow water analogy for a ballistic field effect transistor: New mechanism of plasma wave generation by dc current," *Phys. Rev. Lett.*, vol. 71, pp. 2465–2468, 1993.
[4] M. Nakayama, "Theory of surface waves coupled to surface carriers," *J. Phys. Soc. Jpn.*, vol. 36, pp. 393–398, 1974.
[5] K. Ogata, *Modern Control Engineering*. Englewood Cliffs, NJ: Prentice-Hall, 1970, pp. 193–198.
[6] S. J. Allen, Jr., D. C. Tsui, and R. A. Logan, "Observation of the two-dimensional plasmon in silicon inversion layers," *Phys. Rev. Lett.*, vol. 38, pp. 980–983, 1977.
[7] D. C. Tsui, E. Gornik, and R. A. Logan, "Far infrared emission from plasma oscillations of Si inversion layers," *Solid State Commun*, vol. 35, pp. 875–877, 1980.
[8] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*. Philadelphia, PA: Sounders College, 1976.
[9] H. J. De Los Santos, *Introduction to Microelectromechanical (MEM) Microwave Systems*. Norwood, MA: Artech House, 1999.
[10] J. M. Pond, J. H. Claassen, and W. L. Carter, "Measurements and modeling of kinetic inductance microstrip delay lines," *IEEE Trans. Microw. Theory. Tech.*, vol. MTT-35, no. 12, pp. 1256–1262, Dec. 1987.
[11] M. Dyakonov and M. Shur, "Current instability and plasma waves generation in ungated two-dimensional electron layers," *Appl. Phys. Lett.*, vol. 87, pp. 111501-1–111501-3, 2005.
[12] Y. Cao and D. Jena, "Ultrathin AlN/GaN heterojunctions by MBE for THz applications," in *Proc. Mater. Res. Soc. Symp. Proc.*, vol. 955, 2007, Materials Research Society 0955-I13–05.
[13] M. S. Shur and V. Ryzhii, "Plasma wave electronics," *Int. J. High Speed Electron. Syst.*, vol. 13, pp. 575–600, 2003.
[14] W. Knap, F. Teppe, Y. Meziani, N. Dyakonova, J. Lusakowski, F. Boeuf, T. Skotnicki, D. Maude, S. Rumyantsev, and M. S. Shur, "Plasma wave detection of sub-terahertz and terahertz radiation by silicon field-effect transistors," *App. Phys. Lett.*, vol. 85, no. 4, pp. 675–677, 26 Jul. 2004.
[15] N. Dyakonova, A. El Fatimy, J. Łusakowski, W. Knap, M. I. Dyakonov, M.-A. Poisson, E. Morvan, S. Bollaert, A. Shchepetov, Y. Roelens, Ch. Gaquiere, D. Theron, and A. Cappy, "Room-temperature terahertz emission from nanometer field-effect transistors," *Appl. Phys. Lett.*, vol. 88, pp. 141906-1–141906-3, 2006.
[16] A. E. Fatimy, "Plasma oscillations in nanotransistors for room temperature detection and emission of terahertz radiation," *Phys. Stat. Sol. (c)*, vol. 5, no. 1, pp. 244–248, 2008.
[17] A. K. Rajagopal, G. S. Grest, and J. Ruvals, "Theory of plasmon interactions in an electron gas: Scattering of two plasmons," *Phys Rev. B*, vol. 14, no. 1, pp. 67–75, 1 Jul. 1976.
[18] S. S. Mandal and D. N. Tripathy, "Equation of motion approach to the plasmon-plasmon interaction in a many-electron system," *Phys. Letts.*, vol. 72A, no. 6, pp. 459–463, 6 Aug. 1979.
[19] H. Taub and D. Schilling, *Digital Integrated Electronics*. New York: McGraw-Hill, 1977.

[20] H. J. De Los Santos, *Principles and Applications of NanoMEMS Physics*. Dordrecht, The Netherlands: Springer-Verlag, 2005.
[21] J. R. Krenn, J. C. Weeber, A. Dereux, E. Bourillot, J. P. Goudonnet, B. Schider, A. Leitner, F. R. Aussenegg, and C. Girard, "Direct observation of localized surface plasmon coupling," *Phys. Rev. B*, vol. 60, no. 7, pp. 5029–5033, 15 Aug. 1999.
[22] H. Raether, *Surface Plasmons on Smooth and Rough Surfaces and on Gratings,* vol. 111. New York: Springer-Verlag, 1988.
[23] C. H. van Berkel, M. B. Josephs, and S. M. Nowick, "Scanning the technology: Applications of asynchronous circuits," *Proc. IEEE*, vol. 87, no. 2, pp. 223–233, Feb. 1999.
[24] R. D. Mattuck, *A Guide to Feynman Diagrams in the Many-Body Problem*, 2nd ed. New York: Dover, 1992.
[25] A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinskii, *Methods of Quantum Field Theory in Statistical Physics*. New York: Dover, 1963.
[26] D. Pines, *Elementary Excitations in Solids*. New York: W. A. Benjamin, 1963.
[27] D. Pines and P. Nozières, *The Theory of Quantum Liquids,* vol. 1. New York: W. A. Benjamin, 1966.
[28] C. Itzykson and J.-B. Zuber, *Quantum Field Theory*, International ed. New York: McGraw-Hill, 1985.
[29] D. Tong. (2008). Lectures on quantum field theory [Online]. Available: http://www.damtp.cam.ac.uk/user/tong/qft.html
[30] E. Gerjuoy, "On Newton's third law and the conservation of momentum," *Amer. J. Phys.*, vol. 17, no. 8, pp. 477–482, 1949.
[31] M. Dyakunov and M. Shur, "Detection, mixing, and frequency multiplication of terahertz radiation by two-dimensional electronic fluid," *IEEE. Trans. Electron Devices*, vol. 43, no. 3, pp. 380–387, Mar. 1996.
[32] S. E. Kocabas, G. Veronis, D. A. B. Miller, and S. Fan, "Transmission line and equivalent circuit models for plasmonic waveguide components," *IEEE J. Sel. Top. Quantum Electron.*, vol. 14, no. 6, pp. 1462–1471, Nov./Dec. 2008.

Héctor J. De Los Santos (S'78–M'88–SM'95–F'06) was born in Santo Domingo, Dominican Republic. He received the B.Sc. degree in electrical engineering from the University of Puerto Rico, Mayagüez, in 1979, the M.Sc. degree in engineering from the University of California, Los Angeles (UCLA), in 1981, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, in 1989.

He is currently the President and the Chief Technical Officer of NanoMEMS Research, Limited Liability Company, Irvine, CA, where he is engaged in research on discovering fundamentally new devices, circuits, and design techniques to implement nanoelectromechanical quantum circuits and systems and RF MEMS (NanoMEMS) systems-on-chip. Prior to founding NanoMEMS in 2002, he spent two years as a Principal Scientist at Coventor, Inc., Irvine, CA, where he led Coventor's intellectual property R&D effort, with activities including the conception, modeling, and design of novel radio frequency microelectromechanical system (RF MEMS) devices. From 1989 to 2000, he was with the Hughes Space and Communications Company, Los Angeles, CA, as the Principal Investigator and the Director of the Future Enabling Technologies IR&D Program. Under this program he pursued research in the areas of RF MEMS, quantum functional devices and circuits, and photonic bandgap devices and circuits. He is the author of bestseller textbooks, including *Introduction to Microelectromechanical (MEM) Microwave Systems* (Norwood, MA: Artech House, 1999), *RF MEMS Circuit Design for Wireless Communications* (Norwood, MA: Artech House, 2001), and *Principles and Applications of NanoMEMS Physics* (Dordrecht, The Netherlands: Springer, 2005). He is the holder of over 20 U.S. and European patents. His research interests include, theory, modeling, simulation, design and applications of RF MEMS, semiconductor devices, photonic crystals, plasmonics, and mechanical systems in the quantum regime and nanoelectromechanical quantum circuits and systems (NEMX).

Dr. De Los Santos is a member of Tau Beta Pi, Eta Kappa Nu, and Sigma Xi. From 2001–2003, he lectured worldwide as an IEEE Distinguished Lecturer of the Microwave Theory and Techniques Society. Since 2006, he has been an IEEE Distinguished Lecturer of the Electron Devices Society. The German Research Foundation (DFG) has awarded him a Mercator Guest Professorship to spend the 2010–2011 academic year at the Institut für Hochfrequenztechnik und Elektronik (IHE), Karlsruher Institut für Technologie (KIT), where his activities will include conducting research and developing and teaching courses in the area of NEMX.