A Regulated Charge Pump for Injecting Floating-Gate Transistors

Mir Mohammad Navidi and David W. Graham
Lane Department of Computer Science and Electrical Engineering
West Virginia University, Morgantown, WV 26506
Email: minavidi@mix.wvu.edu and david.graham@mail.wvu.edu

***Abstract*—Floating-gate transistors have found many applications in both analog and digital systems due to their ability to serve as long-term non-volatile memory. When used in programmable analog applications, the charge on the floating gates must be set very precisely, thereby necessitating precise and steady high voltages to generate programming conditions. In this work, we present a regulated charge pump with extremely low output ripple (<1mV) that can be used for accurate programming of floating-gate transistors with channel hot-electron injection. To reduce the ripple to less than 1mV, we present a method to include a low drop-out regulator inside the charge-pump regulation loop. This $0.1196mm^2$ charge pump was fabricated in a 0.35μm standard CMOS process. While operating from a 2.5V supply, this charge pump generates regulated voltages up to 10V.**

I. Introduction

Floating-gate (FG) transistors are one of the most common forms of solid-state non-volatile memory cells. These CMOS-compatible devices are the basis of Flash memories [1] and can also be used to design compact, low-power, programmable analog parameters [2]. Floating-gate transistors store information in the form of trapped charge on an electrically floating polysilicon gate. Under nominal operating voltages, this charge will remain fixed on the gate. Fowler-Nordheim tunneling and channel hot-electron injection are used to program specific amounts of charge onto the gate using high voltages by enabling electrons to pass through the gate oxide.

In [3], we presented a charge pump capable of generating the voltages required for programming, and we focused on tunneling which is the larger voltage that must be generated. We were able to show that the charge pump was able to produce high voltages with relatively low ripple in the output voltage. Tunneling is reserved for global erasure in many analog FG applications [4], and so a moderate amount of ripple on the tunneling voltage is acceptable. However, injection is often used for precise programming, and thus the output ripple should be very small (on the order of 1mV or less).

In this paper, we extend our previous results in [3] to achieve extremely low ripple in the output voltage to provide voltages for injection. We present a method to reduce output ripple in high-voltage charge pumps by adding a high power-supply-rejection-ratio (PSRR) low drop-out regulator (LDO) in the regulation loop of the charge pump. This $260\mu m \times 460\mu m$ charge pump was fabricated in a standard 0.35μm CMOS process. While operating from a 2.5V supply, the charge pump generates regulated voltages from 3V to 10V. The maximum efficiency of the charge pump is 23% for 25μA of load current.

This material is based upon work supported by the National Science Foundation under Award No. 1148815.

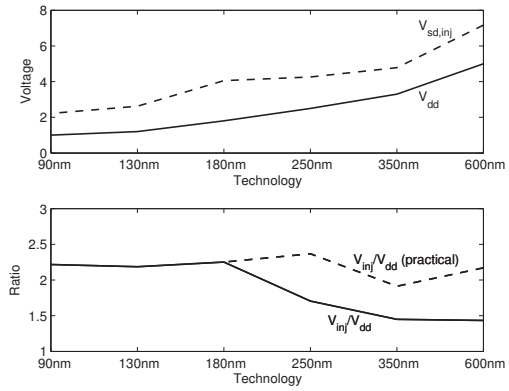

Fig. 1. (Top) Scaling of critical programming voltages: the core supply voltage (V_{dd}) and the write (i.e. injection) voltage ($V_{sd,inj}$), (Bottom) Ratio of the write ($V_{sd,inj}$) voltage to the core supply voltage.

The output ripple is less than 1mV for a wide range of load currents and output voltages. We describe the development of this regulated charge pump in the remainder of this paper.

II. Considerations Regarding Injection Voltages and Scaling

Channel hot-electron injection is used to precisely set a specific charge on an FG in many programmable analog systems [5]-[6]. The injection currents are functions of both the process parameters and the voltages applied to the FG transistor, as given by [7]:

$$I_{inj} = \gamma I_d \frac{V_{gd} + V_{TP}}{0.22 t_{ox}^{1/3} x_j^{1/2}} \exp\left(-\frac{\delta 0.22 t_{ox}^{1/3} x_j^{1/2}}{V_{gd} + V_{TP}}\right) \quad (1)$$

where $\gamma = 3$ and $\delta = 4.9 \times 10^8$ are fits that we have extracted across multiple processes and device sizes. Figure 1 shows how the voltages needed for injection ($V_{sd,inj}$) scale with technology nodes. Despite the stagnant oxide thickness, $V_{sd,inj}$ continues to scale because of X_j becomes smaller for smaller technologies.

Charge pump circuits are used to multiply the chip supply voltage by a constant and generate the voltage required for programming. A charge pump used for programming is enabled for a short period of time to minimize the energy consumption. The input to the charge pump is the chip supply voltage. The step-up ratio of the charge pump that is required for injection is a technology-dependent parameter and is shown in Fig. 1 (bottom). The step-up ratio is lowest for the 250nm through 600nm nodes because below the 250nm node, higher-voltage I/O devices are used to make low-leakage floating

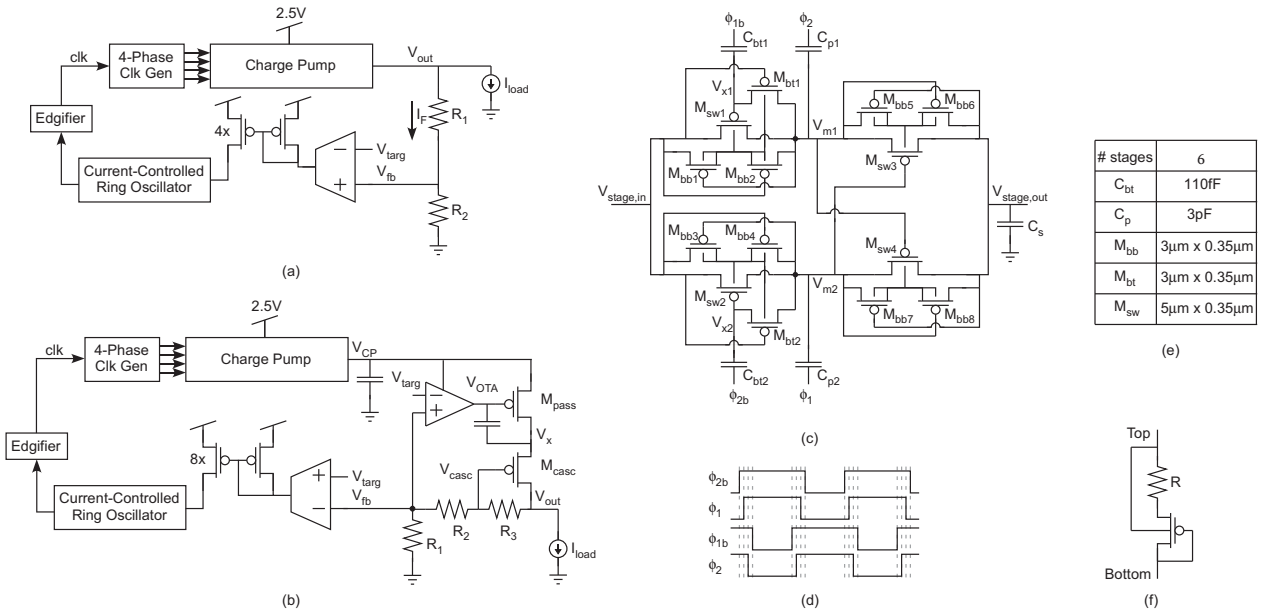

Fig. 2. (a) Block diagram of tunneling charge pump we presented in [3]. (b) Block diagram of the proposed injection charge pump. (c) Circuit diagram of the charge transfer switch (CTS). (d) Non-overlapping clock signals necessary for the CTS. (e) Charge pump specifications. (f) Transistor-level details of the resistor cell to implement the resistive divider.

gates. However, the rated V_{dd} of a technology node causes too much hot-electron injection in standard thin-oxide transistors for stable floating-gate operation. Consequently, the 250nm through 600nm nodes must use lower supply voltages in practice, so the ratios for these nodes are essentially one integer value higher than shown in the figure. Since the programming infrastructure [8] requires an additional overhead of several hundred millivolts beyond the necessary V_{sd} for injection, a charge pump for injection must be able to provide $(2 \sim 3) \times V_{dd}$. In addition, in order to have a suitable charge pump for the injection process, we must have a circuit that has low output ripple. Some other aspects of a good charge pump are fast start-up time and small die area.

III. Proposed High Voltage Charge Pump

Figure 2(a) shows the block diagram of the charge pump we presented in [3]. A voltage divider shifts the output voltage to a voltage between ground and V_{dd}. The difference between this feedback voltage (V_{fb}) and the target voltage (V_{targ}) is converted into a current. This current modulates the frequency of the current-controlled ring oscillator. The open-loop charge pump used in this work includes 6 stages which generate a high-voltage that is related to the oscillation frequency of the oscillator. The circuit schematic of the charge transfer stage (CTS) is shown in Fig. 2(c) [3]. Figure 2(b) shows the circuit diagram of the proposed high-voltage charge pump. The charge pump presented in [3] has a small ripple (around 20mV) which is adjustable with the load capacitor. In this work, we used an LDO inside the loop of [3] to make the ripple smaller. We did not place the LDO outside the charge pump loop because that resulted in larger ripple and longer settling time.

The operation of the closed-loop system of Fig. 2(b) is as follows. Starting from V_{out}: V_{out} is divided by r = 5, then it is converted to a current with transconductance G_m, a current mirror scales this current by a factor of 8, the current-controlled oscillator then converts this current to a frequency with a gain of $K_{RO} = 2kHz/nA$, and finally, the charge pump converts this frequency to the output voltage of the charge pump (V_{CP}). The design specifications of the charge transfer switches used in this work are provided in Fig. 2(e). V_{CP} has big ripple which makes it unsuitable for precise programming; however it is used as the supply voltage of the low drop-out regulator (LDO). The sizes of the resistors in the LDO voltage divider are set to: $R_1 = R_2/3 = R_3$. Since the voltage at the negative input of the operational transconductance amplifier (OTA) inside the LDO loop is at V_{targ}, the LDO will regulate the output voltage to be $5V_{targ}$.

The pass transistor used in this charge pump is a series connection of two pMOS transistors. This structure is similar to the high-PSRR LDO presented by [9]. This cascode structure improves the output impedance of the pass transistor, which leads to an improved supply rejection. As mentioned in [9], this structure will reduce the gate capacitance associated with the pass transistor and will make compensation of the LDO easier. Thick-oxide devices are used for M_{pass} and M_{casc}, since the voltage across these transistors can be large and may damage the devices. Using the saturation region condition for pMOS transistors and knowing that the voltage at the gate of the M_{casc} is at $V_{casc} = 4V_{out}/5$, the maximum voltage allowed at the output to keep M_{casc} in saturation region is:

$$5|V_{Th,thick}| > V_{out} \tag{2}$$

where $V_{Th,thick}$ refers to the threshold voltage of the thick-oxide transistors. Since the body of M_{casc} is connected to V_{CP}, the voltage difference between the body and source of this transistor (V_{bs}) will increase the threshold voltage of M_{casc}. Hence, the upper limit ($5V_{Th,thick}$) on the output voltage increases, too. The range of voltages used for the injection process in 0.35μm CMOS is between 5V and 6.5V; this upper limit ($5V_{Th,thick}$) is high enough to cover the injection voltage range. The saturation region condition for the pass transistor (M_{pass}) can be expressed by:

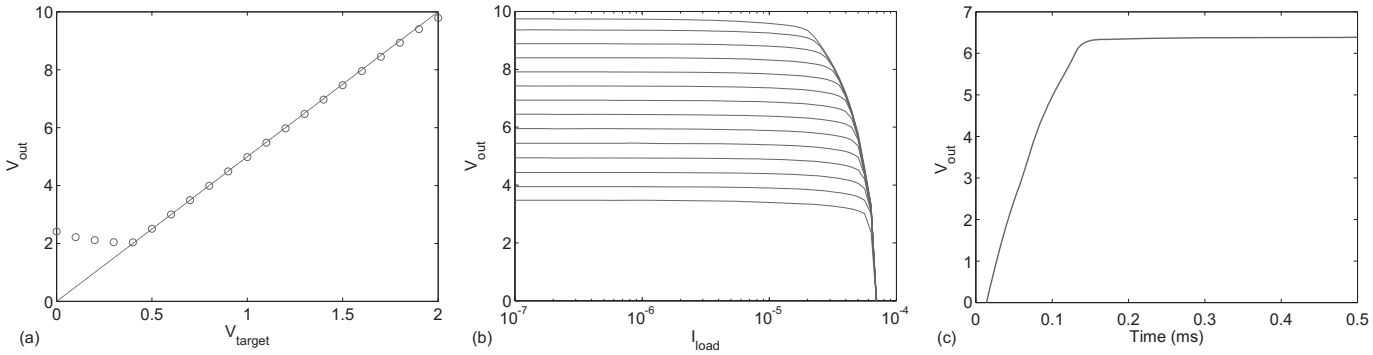

Fig. 3. (a) Measured DC dependence of the charge-pump output on V_{targ} for a purely capacitive load. (b) Measured load regulation of the closed-loop charge pump for various output voltages (c) Measured transient response of the closed-loop charge pump for a 20μA load current.

$$V_{CP} > \frac{4V_{out}}{5} + V_{Th,casc} + \sqrt{\frac{2I_{load}}{K_{pass}}} + \sqrt{\frac{2I_{load}}{K_{casc}}} \quad (3)$$

where $K_i = \mu C_{ox}(W/L)_i$ is the aspect ratio of M_i, $V_{Th,casc}$ is the threshold voltage of the cascode transistor, and I_{load} is the load current.

According to (3), for a specific output voltage (V_{out}) and load current (I_{load}), M_{pass} can be biased in the saturation region by correct selection of K_{pass}, K_{casc}, and the open-loop charge-pump parameters to set V_{CP}. The output voltage of the open-loop charge pump can be set by the following equation [10]:

$$V_{CP} = (N+1)V_{dd} - N\frac{I_{load}}{C_P f} \quad (4)$$

To make V_{CP} satisfy (3), the number of charge transfer stages (N) and the size of the pumping capacitors (C_p) must be adjusted correctly. The operational frequency of the charge pump (f) is a function of I_{load} and V_{out} and will be adjusted by the regulating operation of the charge pump.

Figure 2(e) shows a single resistive cell used in the voltage divider branch of the LDO. Each cell includes a resistor ($98k\Omega$) in series with a pFET, all in a single well, to reduce the overall size of the resistor divider circuitry. R_1 and R_3 are each a single resistive element, and R_2 is a series of three of the resistive elements as shown in Fig. 2(f). The resistance seen between the top and bottom terminals of this resistive cell can be expressed as:

$$R_o = \frac{(g_m + g_{mb})R + \frac{R}{r_o} + 1}{g_m + \frac{1}{r_o}} \quad (5)$$

IV. PSRR OF LDO

In this Section, we establish a frequency-domain equation for the PSRR of the proposed LDO based on [11]. When analyzing only the LDO of Fig. 2(a), a small-signal input voltage, V_i, will induce an output voltage, V_o. The ratio between V_o and V_i simply represents the LDO PSRR. The PSRR of the proposed LDO can be expressed as (7) at the bottom of this page, where Z_L can be expressed as the parallel combination of load capacitance (C_L) and the total resistance of the resistor divider ($5R_o$):

$$Z_L = \frac{5R_o}{1 + 5R_o C_L s} \quad (6)$$

Assuming that $(g_{m1},\ g_{m2},\ g_{mb}) \ll (g_{ds1},\ g_{ds2})$ and that the resistance seen by the resistor divider $(5R_o)$ is high, the PSRR at DC can be approximated as:

$$\frac{V_o}{V_i}|_{DC} = \frac{1}{\frac{g_{m1}A_{ota}}{5g_{ds1}} + 1} \quad (8)$$

Hence, by having a high-gain OTA (A_{OTA}) and keeping M_{pass} in saturation, the denominator of (8) will have a large value, thus significantly improving the PSRR.

V. MEASUREMENT RESULTS

This charge pump was fabricated in a standard n-well 0.35μm CMOS process, and the die area of the charge pump is $260\mu m \times 460\mu m$. A conventional 5 transistor differential amplifier was employed for both error amplifiers. The OTA used inside the charge-pump loop has a pFET input structure, and the OTA used inside the LDO loop has an nFET input structure. The bias currents of the OTAs inside the charge-pump loop and the LDO loop are 270nA and 4μA, respectively. The measured transfer curve from V_{targ} to V_{out} is shown in Fig. 3(a). Deviation at the high voltages is caused by the maximum limit that is provided by (2) and (3). Deviation at the low voltages in Fig. 3(a) is caused by the error-amplification OTA's input transistors in the LDO being pushed out of the saturation region. The closed-loop load regulation of the charge pump is shown in Fig. 3(b). The injection process requires that the output impedance be designed such that the output of the charge pump stays constant in a wide range of load currents. Hot-electron injection requires voltages higher than V_{dd} to make electrons pass through the gate oxide. For our

$$\frac{V_o}{V_i} = \frac{5(g_{m2}g_{ds1} + g_{m2}g_{m1} + g_{mb}g_{m1} + g_{ds2}g_{ds1} + g_{m1}g_{ds1})(1 + R_a C_a s)}{(g_{m2} + g_{mb} + g_{ds2})g_{m1}A_{ota} + \frac{5(g_{m2}+g_{mb}+g_{ds2})(1+R_a C_a s)}{z_L} + (4g_{m2}g_{ds1} + 5g_{ds2}g_{ds1})(1 + R_a C_a s)} \quad (7)$$

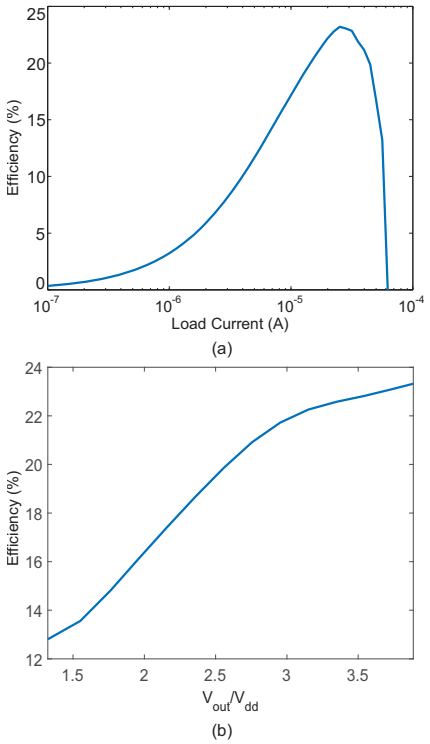

Fig. 4. (a) Measured efficiency of the charge pump versus load current when $V_{out} = 6.5V$ (b) Measured efficiency of the charge pump versus V_{out}/V_{dd} for 20μA of load current.

application, we will mostly use 6.5V as the elevated supply voltage in the injection process.

Figure 3(c) shows the measurement results of the charge pump generating a 6.5V injection pulse with 20μA of load current for a load capacitance of 0.47μF. This plot shows that the output voltage ramps up to the target voltage in less than 0.2ms.

The efficiency of a charge pump can be expressed by:

$$\gamma = \frac{V_{out} I_{load}}{V_{dd} I_{vdd}} \tag{9}$$

which is the power delivered to the load divided by the total power going into the charge-pump circuit. Figure 4(a) shows the efficiency measurement of this charge pump for 6.5V output. The maximum efficiency is around 23%, and it starts to drop for load currents greater than 25μA because the loop gain is not high enough to hold the output voltage at 6.5V. Figure 4(b) shows the efficiency measurement of this charge pump for different step-up ratios (V_{out}/V_{dd}) for 20μA of load current. This plot shows that this charge pump is able to provide higher voltages with better efficiency.

The accuracy of the injection process is a function of the accuracy of the elevated injection voltage (e.g. 6.5V in our application). The required accuracy for programming will set a maximum limit on the ripple voltage (on the order of 1mV or less). Because the ripple was very small and within the limits of our measurement equipments, we used RC extracted simulations to report the ripple amplitude. Figure 5 shows the RC-extracted simulation results when the charge pump was set to 6.5V output. The load current was set to 0μA and 20μA in these simulations, and the simulated peak-to-peak output ripple is approximately 0.4mV and 0.6mV, respectively.

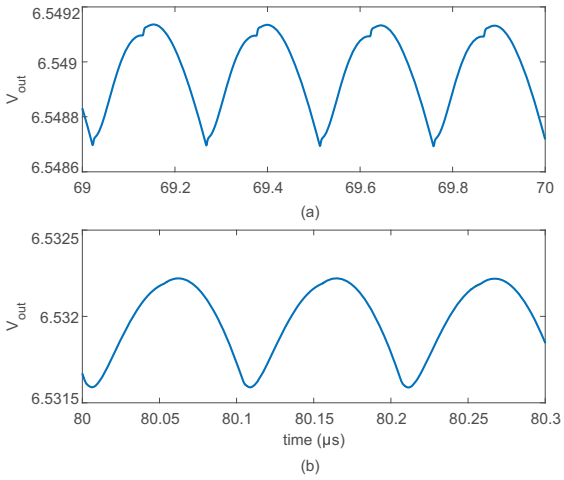

Fig. 5. Simulated ripple for an output voltage of 6.5V with (a) 0μA load current and (b) 20μA load current.

VI. Conclusion

In this paper, the design of a regulated charge pump for injection of FG transistors is presented. A high PSRR LDO is used inside the charge-pump loop to improve the output impedance of the charge-pump and the ripple voltage to less than 1mV. This charge pump has a compact size and consumes low power which makes it suitable for battery-powered applications.

References

[1] K.-D. Suh, *et al.*, "A 3.3 v 32 Mb NAND flash memory with incremental step pulse programming scheme," *IEEE Journal of Solid-State Circuits*, vol. 30, no. 11, pp. 1149–1156, Nov. 1995.

[2] B. Rumberg and D. W. Graham, "A low-power and high-precision programmable analog filter bank," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 59, no. 4, pp. 234–238, April 2012.

[3] B. Rumberg, D. W. Graham, and M. M. Navidi, "A regulated charge pump for tunneling floating-gate transistors," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 64, no. 3, pp. 516–527, March 2017.

[4] P. D. Smith, M. Kucic, and P. Hasler, "Accurate programming of analog floating-gate arrays," in *IEEE International Symposium on Circuits and Systems*, vol. 5, 2002, pp. V–489–V–492.

[5] S. Shapero, C. Rozell, A. Balavoine, and P. Hasler, "A scalable implementation of sparse approximation on a field programmable analog array," in *BioCAS*, Nov. 2011, pp. 141–144.

[6] B. Rumberg, D. W. Graham, S. Clites, B. M. Kelly, M. M. Navidi, A. Dilello, and V. Kulathumani, "RAMP: Accelerating wireless sensor hardware design with a reconfigurable analog/mixed-signal platform," in *IPSN*, 2015, pp. 47–58.

[7] T.-C. Ong, P.-K. Ko, and C. Hu, "Hot-carrier current modeling and device degradation in surface-channel p-MOSFETs," *IEEE Transactions on Electron Devices*, vol. 37, no. 7, pp. 1658–1666, July 1990.

[8] B. Rumberg and D. W. Graham, "A floating-gate memory cell for continuous-time programming," in *Proceedings of the Midwest Symposium on Circuits and Systems*, Aug 2012, pp. 214–217.

[9] K. Wong and D. Evans, "A 150mA low noise, high PSRR low-dropout linear regulator in 0.13μm technology for RF SoC applications," in *European Solid-State Circuits Conference*, Sept. 2006, pp. 532–535.

[10] G. Palumbo and D. Pappalardo, "Charge pump circuits: An overview on design strategies and topologies," *IEEE Circuits and Systems Magazine*, vol. 10, no. 1, pp. 31–45, 2010.

[11] A. Amer and E. Sanchez-Sinencio, "A 140mA 90nm CMOS low drop-out regulator with -56db power supply rejection at 10MHz," in *Proceeding of Custom Integrated Circuits Conference*, Sept. 2010, pp. 1–4.