Low Power Precision Current Distribution Network

Piotr Mitros

under the direction of
Professor Rahul Sarpeshkar
Neuromorphic VLSI Lab

December 15, 2000

Abstract

This paper presents a method for creating and distributing small, precision currents throughout an integrated circuit. The system is designed for use in a bionic cochlear implant. A bandgap reference and an off-chip precision resistor create a single precision reference current. This current is then converted into an oscillation frequency. This oscillation frequency is distributed losslessly around the chip, and is converted back to current where it is needed. By clocking down the frequency of oscillation, the magnitude of the current can be varied by several orders of magnitude.

Contents

List of Figures

1   Introduction

This paper presents an unconventional technique for generating bias currents from the low microamp down to the picoamp range for use in ultra-low power designs. Generating these currents in a conventional manner would require very large resistors, beyond the range of what is possible in a typical process. As a result, most subthershold designs either rely on very crude bias current levels, or require significant numbers of external components. Our technique overcomes these limitations, although at a high cost in complexity, and a minor cost in power utilization. This technique is implemented as a circuit designed for use in a bionic cochlear hearing implant, which must be powered for years from a small battery. However, it is applicable to most other subthreshold very large scale integration (VLSI) designs as well.

The first step in this technique is to generate a reference current. The second step is to convert this current into an oscillation frequency, relative to a reference voltage and capacitance. The third step is to clock down this frequency. The final step is to convert the new frequency back to a current, given an identical reference voltage and capacitance. This new current is the output of the circuit.

The initial reference current is set by a bandgap reference and a precision off-chip resistor. The frequency-to-current conversion is done using a linear current-controlled oscillator based on a spiking neuron circuit. Charge is dumped onto a capacitor until the capacitor voltage reaches a reference level. Once this level is reached, the capacitor is completely discharged. Clocked-down frequencies are converted back to current by frequency-locking them to an identical spiking neuron circuit. The input current to this neuron is $I = I_0 \cdot \frac{f}{f_0}$. The current entering the spiking neuron circuit can be extracted with a current mirror. For frequencies that do not lie precisely on powers of two of the initial reference current, the $\frac{W}{L}$ ratio on this mirror can be changed.

This design was chosen because a standard process is capable of producing the following devices:

- Crude resistors[1]

- Crude absolute values of capacitors ($\pm 20\%$)

- Precision matching of capacitors ($0.1\%/\sqrt{pf}$ matching)

- Precision voltage sources (using a bandgap reference)

Each of the above gives a unit (precise voltage, precise ratios, crude resistance and capacitance). By dimensional analysis, some off-chip element is needed to create precise units of current. As a result, the initial reference current is created using a single off-chip precision $1M\Omega$ resistor. A voltage from a bandgap reference is run across this resistor, thereby creating a single precision reference current, which can then be scaled.

Given a single reference current, scaling this current across multiple orders of magnitude in order to produce bias currents is a difficult problem. Simply cascading current mirrors

---

[1]Although processes exist that allow precision on-chip resistors, they are expensive, and not easily accessible.
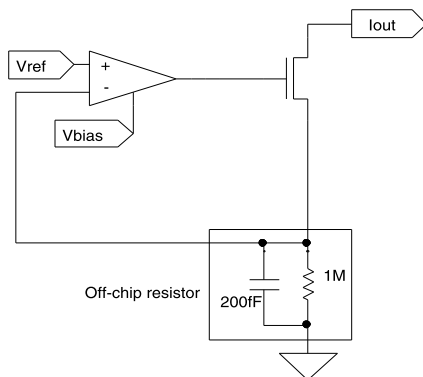
Figure 1: Basic Current Source Design

would accumulate very significant mismatch error. To address this problem, this paper develops a technique based on the perfect integrity maintained by digital signals, the very good matching of capacitors and the consistency of a voltage reference across an integrated circuit.

Furthermore, communicating small currents across an integrated circuit die is difficult. Capacitive coupling from digital signals, power supply noise, and other analog signals can generate significant errors. In addition, radio frequency pickup (especially in a radio-powered application, as found in many subthreshold designs) can contribute additional noise. Finally, parasitic resistance and capacitances may render the signal useless if it is communicated across a large die. Transporting the current magnitude as a digital clock frequency renders it impervious to all of these effects.

2 Precision Current Source

2.1 Basic Design

In order to generate the precision current source, a precise voltage reference is run across an external resistor. The use of a simpler RC relaxation oscillator topology is impossible because of the uncontrolled parasitic capacitance across the leads of the external resistor. These parasitics can be completely ignored by keeping the voltage across the resistor constant.

Buffering the voltage from a bandgap reference holds the voltage across the resistor constant. Unfortunately, a simple operational transconductance amplifier (OTA) buffer, as shown in Figure 1, lacks sufficient gain to hold the voltage constant. At $V_{REF} = 0.5$V, this gives an error of about 5%, varying with power supply voltage. Instead, a network of 3 OTAs is used, in a topology similar to that of an instrumentation amplifier, as shown in Figure 2. This network gives twice the square of the open-loop gain of a single OTA.

In simulation, for a power supply voltage of $V_{DD} = 1.8$V, this circuit network buffers 0.5V to 0.4999918V, and at $V_{DD} = 3.8$V, $V_{REF} = 0.5$V buffers to $V_R = 0.4999926$V, as shown in Figure 3. Therefore, this error will be dominated by transistor mismatches, parasitics and
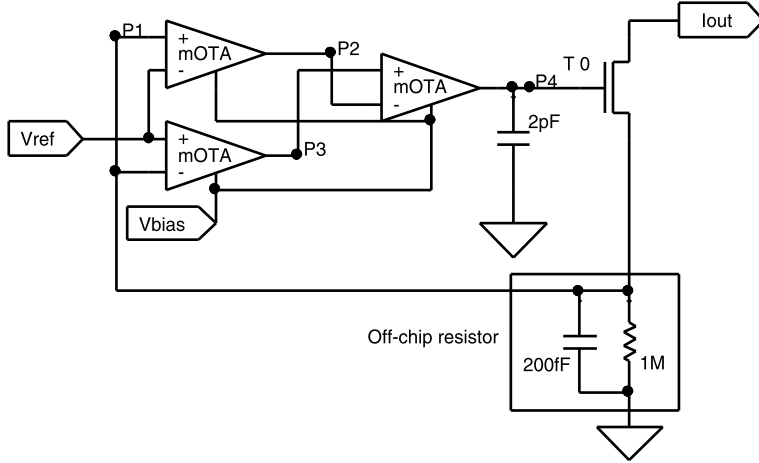
Figure 2: Final Current Source Design

other divergences from ideal behavior.

2.2 Offset

Due to the exponential V_{GS}-current relationship of the MOSFET in the subthreshold region, and the lack of precision in control over the threshold voltage of the transistors, transistor mismatch normally contributes significant amounts of error to subthreshold designs. In the case of an OTA, or other operational amplifier topologies, the transistor mismatch manifests itself as an input offset voltage. The topology presented is designed to minimize error due to input offset. Offset contributed by the output OTA is divided by gain, and so is negligible. The offset of the input OTAs is:

$$O_{TOTAL} \approx \frac{g_1 \cdot O_1 - g_2 \cdot O_2}{g_1 + g_2} \tag{1}$$

Where g_1 and g_2 are the gains of the two OTAs. If the OTAs have similar gains and offsets, the offsets will tend to cancel.

2.3 Power Usage

Current through the output transistor and resistor dominates the power usage; in comparison, the OTAs use negligible power (due to the large gain of the configuration, the system can maintain reasonable gain with the OTAs biased at very small current levels). Therefore, the total power for the current source is approximately $0.5\mu A \cdot V_{DD}$, and twice that with the current mirror.

With this design, nearly half of the power is wasted in mirroring the output current. Power usage could be cut almost in half, while increasing accuracy of the current, by driving a P-channel version of the current-controlled oscillator directly from the precision current
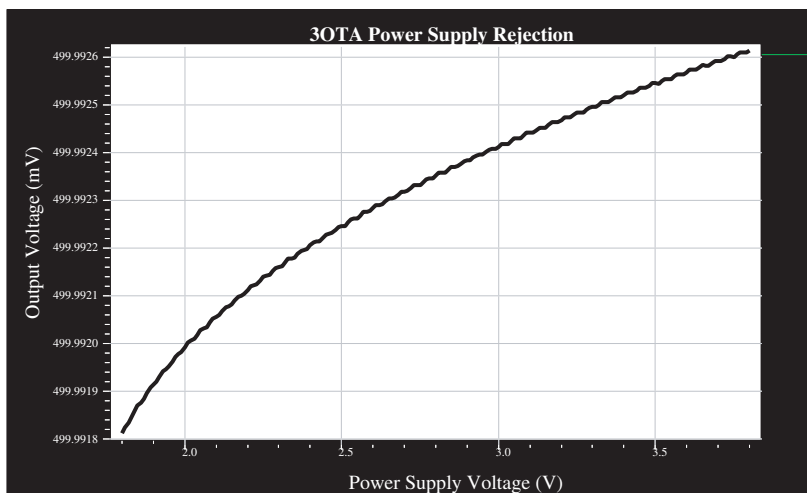
3

Figure 3: 3OTA Current Source Power Supply Rejection

source, rather than mirroring the current and driving an N-channel version as in the current circuit. This change would also eliminate the error introduced by mismatch in the single current mirror. Due to time constraints, this change was not implemented.

2.4 Stability

Since transient performance is of no concern, a simple dominant pole scheme is used to stabilize the design. The dominant pole is injected through a capacitor on the output of the triple-OTA topology, since the output is the most convenient high impedance node. The Miller capacitance is not exploited because, at high frequencies, this capacitance simply couples the input signal over to the output. This coupling injects an additional zero. Although this zero is often ignored in conventional operational amplifiers, in the topology used in this circuit, the high gain of the first stage OTAs would cause this zero to occur well before loop crossover.

One of the secondary poles comes from the parasitic capacitance of the off-chip resistor. Since the capacitance is not very well controlled, some reasonable margin of safety must be left in the dominant pole capacitor. Experimentally, this capacitor was chosen to be 2pf.

3 Oscillator

For the base oscillator, a spiking neuron circuit was used. The original circuit is shown in Figure 4. In this circuit, the input current charges the capacitor. Once the voltage on the capacitor reaches some threshold, the capacitor is rapidly discharged, and the process starts again.

This neuron is not very precise; the voltage at which it spikes is dependent on a current-starved inverter, which in turn is highly susceptible to power-supply fluctuations. Ideally, this
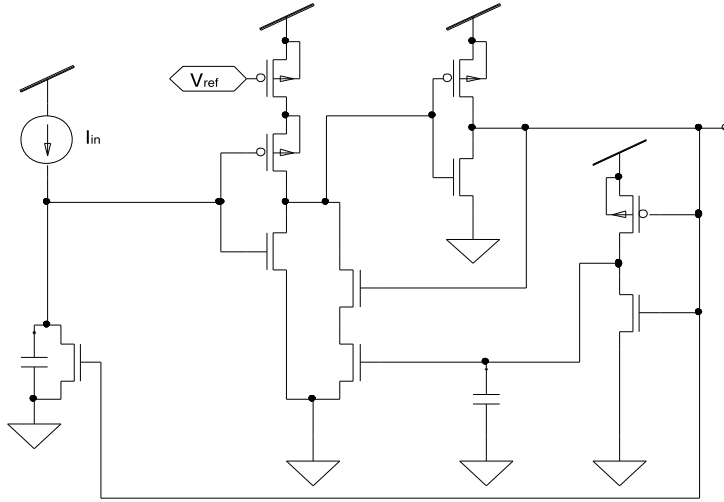
4

Figure 4: Basic Spiking Neuron Circuit

circuit would be replaced with something akin to the circuit in Figure 5. However, building an ultra-low-power operation amplifier with sufficient gain and speed for this application is difficult. A conventional OTA is somewhat slow. As a result, the capacitor is discharged slightly after it passes the reference voltage, rather than at the reference voltage. This delay results in a reduced spiking frequency for a given current level. The amount of error is dependent on the speed of the voltage rise on the capacitor, and therefore on the current level. Since the circuit relies on the precision of the oscillation frequency, this delay would be a major source of error.

Several approaches were used in optimizing the oscillator. Each resulted in a loss elsewhere.

- The bias current on the OTA was increased. The increased current had the disadvantage of significantly increasing power usage, but improved both the speed and the gain of the OTA.

- A small quantity of high-passed positive feedback was run across the OTA. This feedback was implemented as a capacitor connected from the output of the OTA back to the positive input, and connecting the threshold voltage through an adaptive element. The positive feedback sped the OTA up considerably, although not in an entirely well controlled fashion.

- The spiking voltage was increased to 1.2V. As a result, small variations in spike voltage made a proportionally smaller difference in timing. It was not increased further because of power supply limitations. Since the band gap reference does not output 1.2V, this change breaks the precision of mapping from current to frequency. However, in practice, the design only relies on the 1.2V reference being consistent across the chip, not on its absolute value. Therefore, the circuit does not lose precision. A second bandgap reference is not used due to design complexity and die space limitations.
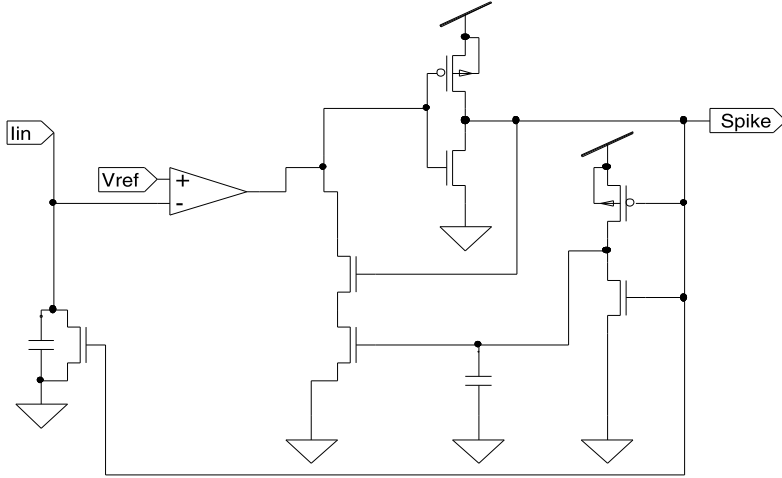
5

Figure 5: Precision-Referenced Spiking Neuron Circuit

- The size of the capacitor was increased to 10pf. This uses additional die space, but also increases charge time, so as to make speed less important.

The final optimized neuron is shown in Figure 6. The circuit was not further optimized, as solving the problem of the tradeoff between amplifier gain, speed and power usage is beyond the scope of this paper. The final version will use a recent, unpublished neuron design from Micah O'Halloran, which uses a metastable latching mechanism to achieve a very precise spike time and position.

4   Counter

The counter consists of a simple cascade of digital flip-flops. To run the input from the spiking neuron circuit into the counter, the saw-tooth output of the neuron is compared to a voltage using a current-starved inverter. If the power usage of the current-starved inverter is P_{INV}, and that of each flip-flop, $P_N \approx \frac{P_1}{2^N}$ (since flip-flops use minimal power unless switching), the total power usage converges to:

$$P_{TOTAL} \approx P_{INV} + P_1 + P_2 + P_3 + ... = P_{INV} + P_1 + \frac{P_1}{2} + \frac{P_1}{4} + ... = P_{INV} + 2 \cdot P_1 \quad (2)$$

As a result, a counter consisting of four levels approximates the total power consumption of a large counter to within less than $1 - \frac{2}{1+0.5+0.25+0.125} \approx 7\%$. In simulation, such a counter, with a neuron spike period of $30\mu S$, used $0.31\mu W$ of power.

An alternative, though inferior, counter architecture would consist of a cascade of spiking neurons. Each would be triggered to spike after it received the amount of charge injected by one and a half of the previous neuron's spikes. That way, each neuron would spike during the previous neuron's second spike. The duration of the output spike would mask out the
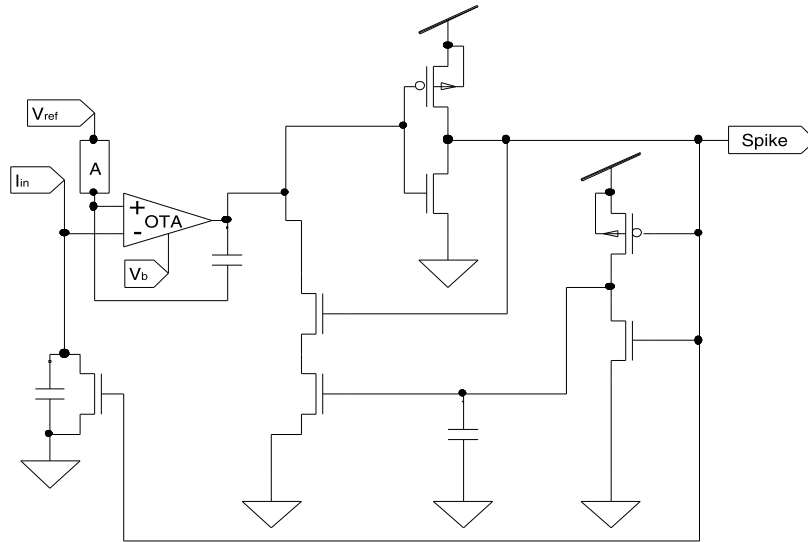
Figure 6: Optimized Spiking Neuron Circuit

remaining charge of the previous neuron's spike that triggered the output spike. While better aligned with the biological goals of the project, this design consumed sufficiently more power, and occupied sufficiently more die area, that the more traditional digital design was chosen.

5 Frequency Lock Loop

A frequency lock loop (FLL) converts the frequency back to a current. The FLL is a less precise, but simpler, variant on the traditional phase locked loop. Instead of measuring a phase difference, the FLL measures a frequency difference. Since the circuit does not rely on an exact phase match for any portion of its operation, this measurement is sufficient. In return, the circuit requires much less calibration, as the lock range in only limited by what the current controlled oscillator can output.

5.1 Frequency Comparator

The frequency comparator circuit compares a count of cycles from each signal. It was based around a floating capacitor design, as shown in Figure 7. Each time one of the input signals had a rising clock edge, a fixed amount of charge was placed on the floating capacitor. Each time the other signal had a rising clock edge, that same amount of charge was taken off of the floating capacitor. As a result, at any point in time, the amount of charge on the floating capacitor reflected the difference in the cycle count between the two oscillators.

The frequency comparator was implemented with a circuit similar to that shown in Figure 8. However, the circuit shown does not work because the voltage on the capacitor tends to drift. Accuracy was degraded through several effects:
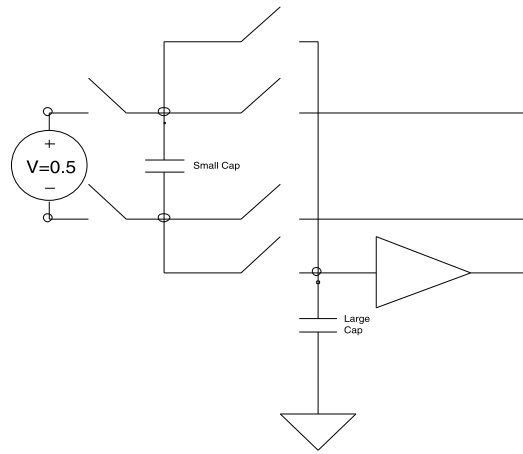
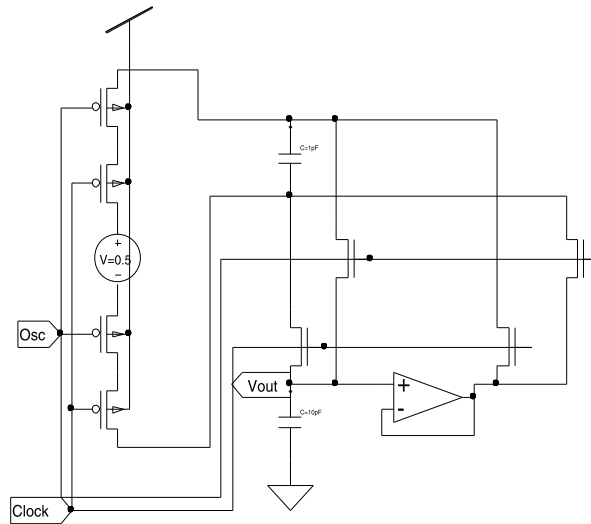Figure 7: Block Diagram of Frequency Comparator

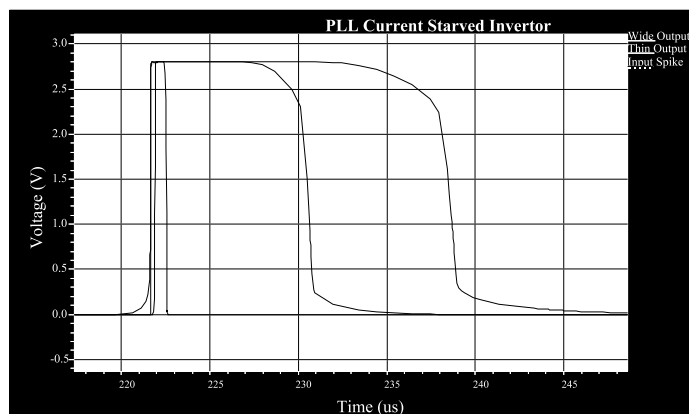Figure 8: Basic Circuit for Frequency Comparator

Figure 9: Current Starved Inverters Enlarge Spikes

- Leakage through the switches when both are partially on

- Parasitic leakages in the large capacitor

- Charge injection

- Buffer offset

- Buffer fluctuations due to load

- Spike collisions

- Lack of time for the small capacitor to fully charge and discharge (depending on the logic driving the switches)

Most of the leakage happened during the switching period when both switches were partially on. This problem was solved by running the input spikes through a pair of current-starved inverters. One was designed to rise immediately, and took a long time to fall. The other rose slowly and took a shorter time to fall. The former drove the switches to the power rails, and the latter drove the switches to the large capacitor. As a result, one switch would have time to fully turn off before the other turned on. The time constants were set such that the small capacitor could almost completely discharge into the large capacitor during the shorter spike. As this time was increased, the probability of spike collisions increased as well, so it was kept as short as possible, with the constraint that the small capacitor still had ample time to discharge. The three spikes are shown in Figure 9.

Some of the charge injection is canceled by switching on both sides of the capacitor. More is canceled by pairing N and P-type transistors as switches, since each injects charge in the opposite direction. Furthermore, the circuit uses very small $(3\lambda \times 4\lambda)$ transistors to minimize gate-source and gate-drain capacitance. The charge injection could be further reduced by using a single switch between the small capacitor and its power line, and adding digital logic to control this switch from the signals currently going into the two switches. However, charge

9

injection was brought within reasonable levels without the added complexity of this digital logic.

In the simplified circuit, the buffer output fluctuated under the load change that occurred during switching. This effect was reduced by placing a large capacitor on the end of the buffer, decreasing its output impedance. This additional capacitor had the side effect that the buffer could not keep up with the input voltage during the discharge; however, the effect of this lag was small and symmetric, and so did not contribute error.

The current circuit has no mechanism for handling spike collisions. An early version of the circuit included digital logic to handle collisions, turning off both switches in the case of a collision. However, in simulations, the addition of this digital logic made little difference, so the logic was removed, for the benefit of simplicity, die area and power consumption.

The circuit generally performed well with respect to collisions, since their effects tended to cancel. However, it occasionally entered a state in which collisions would generate beats. Practically, this state only occurred when the frequencies were very close[2]. In simulation, beats did not occur unless the frequencies compared were within one percent or less of each other. As such, in a worst-case scenario, when it came to within one percent, the beat behavior would induce incorrect behavior. The incorrect behavior would reduce or increase the controlled oscillator frequency, until it fell a percent outside of the controlling frequency. At that point, the frequency-lock loop would regain lock, and begin bringing the frequencies together. This would give a maximum of about one percent of error, which is tolerable. In practice, this error would be significantly lower, since half of the time, the beats would result in a higher frequency, and half of the time, in a lower one.

Combined, the remainder of these effects still tends to push the capacitor gradually towards some fixed voltage around 1.4V. Although this is an undesirable effect, it is possible to minimize its effects through the biasing of the next stage of the circuit. Specifically, if the next stage outputs a reasonable current with an input bias of 1.4V, the loop feedback is only used to compensate for small changes, so the error contributed by this problem becomes negligible.

The final frequency comparator circuit is shown in Figure 10. In simulation, it compared frequencies fairly accurately. The version with a small capacitor (0.1pf and 1pf) could clearly and reliably detect the difference between a $290\mu S$ period and a $300\mu S$ period, or a 3% difference[3]. By $299\mu S$ vs. $300\mu S$, the results were no longer useful. $150\mu S$ vs $300\mu S$ showed a constant difference of over 200 mV[4]. Larger capacitors gave better results, but were not used due to die space limitations.

Power consumption depended on frequency, since almost all power usage occurred during the short spiking periods. With capacitance set to 0.1pf for the small capacitor, 1pf for the large, and a period of around $300\mu S$ (about ten times the original), it used $0.31\mu W$. Power consumption did not change significantly if the capacitors were increased to 1pf and 10pf, as most of the power is used in the current-starved inverters and digital logic.

---

[2]Theoretically, they also appear when the frequencies are far apart, but close to within a multiple of each other, but this state would not come up in practice.

[3]Frequency comparator followed by a low-pass filter with a time constant on the order of $100\mu S$

[4]Intermediate results are not available, as each simulation took about an hour to run
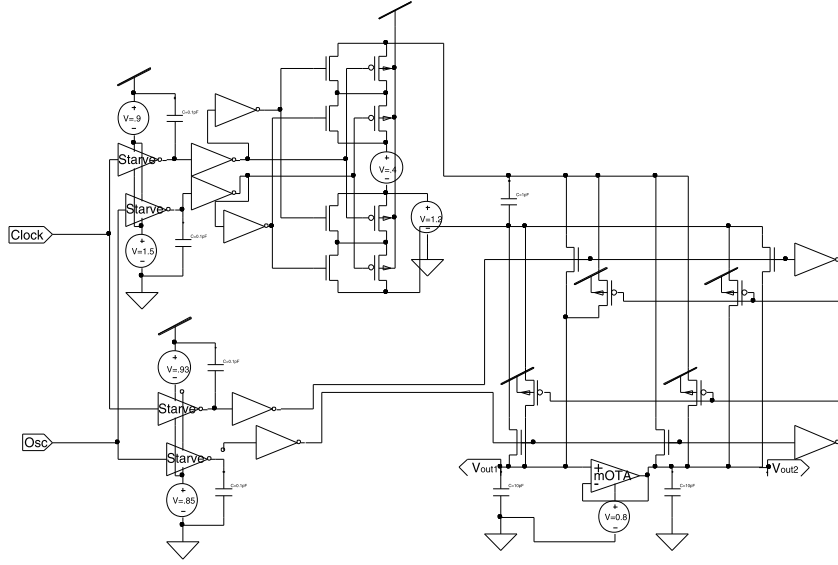
Figure 10: Final Frequency Comparator

5.2   Frequency Lock Loop Compensation

A simple reduced gain scheme compensates the frequency lock loop. The pole comes from the integration of frequency difference in the frequency comparator. Crossover frequency is determined from phase margin from the dominant pole and the phase shift of the delay from the discretization. The other elements contribute no poles or zeros until well after crossover.

The frequency comparator is modeled as an integrator, together with a delay of half a cycle. As such, the open loop transfer function of the frequency-look loop without compensation is:

$$K_D \cdot \frac{K_0}{s} \cdot e^{-\frac{st}{\tau}} \tag{3}$$

where K_D is the gain from current to frequency of the current controlled oscillator. K_0 is the gain of the frequency comparator, from number of pulses to output voltage, and τ is the delay in comparing the frequencies, with:

$$K_D = \frac{f}{I_{in}} = \frac{1}{V_{REF} \cdot C} \tag{4}$$

$$K_0 = \frac{V_{CAP} C_S}{C_L} \tag{5}$$

$$\tau \approx \frac{1}{2f} \tag{6}$$

As a result, there are two obvious compensation schemes:

• Reduced gain compensation.

11

- Integrator, followed by a zero before crossover.

The former is simpler, while the latter would theoretically give lower error. Since lack of accuracy in the frequency lock feedback loop did not contribute significantly to the error, the simpler reduced gain compensation scheme was chosen.

The exact compensation network varies with τ, which varies with the exact output current desired, which in turn varies with the specific context in which the circuit is used. However, the value is easily calculated for the specific application. For a 45° phase margin, crossover occurs at $\frac{\pi}{4\tau}$. This means that our compensator is simply an OTA biased for a gain of:

$$ G = V_{REF} \cdot C \cdot \frac{C_L}{V_{CAP}C_S} \cdot \frac{\pi}{4\tau} $$

Lower gain will not hurt performance, so long as the output current of the OTA can still reach the desired value. The time constant of the OTA is very fast compared to the crossover frequency of the feedback loop. As a result, the poles within the OTA will not effect the feedback loop, even at a very small bias level within the OTA.

The entire frequency lock loop, with compensation, is shown in Figure 11.

6   System Integration

The final system contains several minor modifications beyond simply connecting the stages described above. In order to reduce switching noise from the frequency comparator, the output of the frequency comparator was passed through a low-pass filter (implemented as a current-starved emitter-follower and capacitor). The time constant of this filter was placed well after crossover of the frequency lock loop. The output of the low-pass filter was connected into the input of the compensation element (an OTA with slightly increased linear range). The output of the OTA was current-mirrored into the spiking neuron circuit, in parallel with a voltage-biased transistor current source designed always to err slightly on the low side. The second input of the OTA was biased such that the OTA always gave small positive current within the appropriate range for a 1.4V input.

As described above, this technique still does not completely achieve the ultra-low currents desired. The frequency-locked loop may begin to fail if the frequency drops down to fractions of a Hz, as it is difficult to implement sufficiently slow low-pass filters with the capacitor sizes available in a typical process. Using smaller capacitors in the spiking neuron oscillators scales down the current, potentially by several additional orders of magnitude. Slightly more performance can be achieved by stepping down the final output current with a current mirror. Together, these methods allow for the necessary current levels.

For currents close to that of the initial reference current source, a simple current mirror of the initial source is used, instead of the topology described above. This removes the problem of dealing with high (power-hungry) frequencies beyond the initial oscillator.
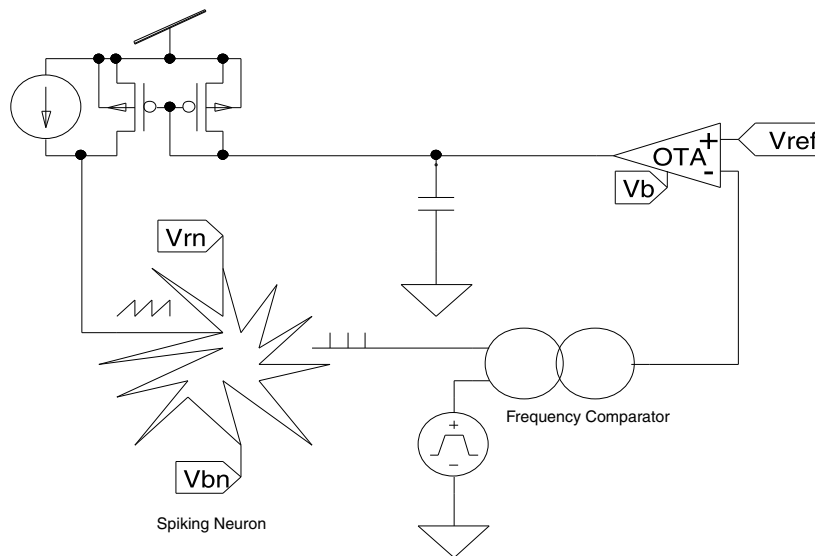
Figure 11: Frequency Locking Loop

7 Layout

The layout of the circuit was completed using L-Edit[5]. Transistors smaller than $8\lambda \times 8\lambda$ appeared only in digital logic portions of the design, or as analog switches. Although all design rules were followed, a portion of the circuit did not pass final design rule check. A portion of the DFLOP in the counter was layed out in a topologically different (although logically equivalent) configuration than in the original circuit, in order to save die space. Similar logical substitutions were made elsewhere where it made sense.

8 Results and Limitations

The components described above were simulated and tested individually. Results for the initial current source were excellent; the error was minimal in simulation, and depending on manufacturing technology, would be limited either by process mismatch of the current mirror, or the precision of the off-chip resistor. The spiking neuron circuit gave a tradeoff between power consumption and error, based on the bias current level of the OTA. For lower power levels, the timing of the spikes was the dominant source of error. The counter operated as expected. The frequency comparator of the frequency lock loop may contribute one additional percent of error, due to the beat effect described.

Unfortunately, system-level simulations were impossible, as the time constants of different portions of the circuit were drastically different. The computational resources required to simulate the system as a whole were unavailable[6].

---

[5]Design of the spiking neuron is in Micah O'Halloran's L-Edit file.

[6]The systems available were limited to $300MHz$ Pentium II machines with $128MB$ RAM each. Longer

In particular, it was impossible to test the frequency-locked loop in a complete feedback configuration, due to the complexity of the spiking neuron circuit, combined with the drastically longer time constant of the frequency lock loop. The frequency comparator takes a moderately large number of cycles to converge, and the frequency locked loop takes several time constants of the frequency comparator. The spiking neuron circuit requires a fairly lengthy simulation during each spiking period. Thus, simulating the entire system would take a minimum of several days of computation time.

The final design also has several known limitations.

8.1   Power Supply Noise

Although total power supply usage of all components was at a fairly low level, due to its high reliance on digital (or at least discretized) computation, this circuit draws power in short, large bursts. The spiking neurons discharge large amounts of current into the ground node. The frequency comparator needs to draw large, short bursts of current to quickly charge the small capacitor. The counter, spiking neurons and many inverters are mostly or purely digital logic. As such, the project may inject large amounts of power supply noise, if the power supply is not very well bypassed externally. Although the circuit has an excellent power supply rejection ratio at DC, its performance with AC high frequency power supply noise has not been characterized.

8.2   Simulation of Switched Capacitors

There were also some minor problems with T-Spice's handling of the high-frequency noise from the frequency comparator. The large transient currents of the switching capacitor circuit would cause T-Spice to require low current noise tolerances. Simulations with more complex filters either took hours to run, or did not converge with tolerances needed for other portions of the circuit.

9   Future Improvements

Current noise from switching of the frequency comparator could be further reduced by adding a fast time scale peak detector after the frequency comparator. This peak detector could eliminate most of the switching noise. The most basic way to implement the peak detector would rely on a traditional crystal detector topology. This topology would be limited in performance, but still achieve significant gains. A more aggressive topology, consisting of an oscillator-triggered sample-and-hold circuit, would virtually eliminate switching noise.

10   Conclusion

The circuit presented generates currents from the microamp range down to the picoamp range for use in a hearing implant. It overcomes a number of limitations of traditional

---

simulations were run in parallel on several machines, with each handling a different parameter range.

techniques for generating similar currents in subthreshold designs. It does not require a large number of external components. It is based on well calibrated components. Therefore, it is insensitive to typical process or thermal variations. This level of stability makes it is well suited for use in physically small applications, where it must operate over several years without calibration.