Generating all Negations

Yuval Filmus

December 22, 2009

Puzzle: *Given as many AND and OR gates as you'd like, but only two NOT gates, build a circuit inverting three inputs.* The rest of this note solves this puzzles, poses a generalization, and proves the uniqueness of the optimal solutions.

We begin on the positive side. We show how to invert $2^n - 1$ inputs x_i using only n NOT gates. Note that those functions creatable using only AND and OR gates are exactly the monotone functions — that is functions f such that $x \leq y$ implies $f(x) \leq f(y)$. Hence we can compute HAM_{n-1}, the highest bit of the hamming weight of the inputs, without using NOT gates. We use one NOT gate to invert HAM_{n-1}. If m_0 and m_1 are monotone then we can now compute the function

$$m_{01} = \begin{cases} m_0 & \mathrm{HAM}_{n-1} = 0 \\ m_1 & \mathrm{HAM}_{n-1} = 1 \end{cases}$$

using the formula

$$m_{01} = (\neg \mathrm{HAM}_{n-1} \wedge m_0) \vee (\mathrm{HAM}_{n-1} \wedge m_1).$$

In particular, we can compute HAM_{n-2} without using NOT gates. Continuing this way, we can compute all of the bits of HAM and their inverses using only n inverters. Finally we show how to compute $\neg x_0$. Let w_k be true when at least k of $x_1, \ldots, x_{2^n-2}$ are true. The functions w_k are monotone and so we can compute the function which is equal to w_k when $\mathrm{HAM} = k$. Yet this function is exactly $\neg x_0$!

The solution to out original puzzle is therefore as follows (x_i are inputs, y_i outputs, t_i^1 auxiliaries, t_i^0 their inversions):

$$\begin{aligned}
t_1^1 &= (x_0 \wedge x_1) \vee (x_0 \wedge x_2) \vee (x_1 \wedge x_2) \\
t_1^0 &= \neg t_1^1 \\
t_0^1 &= (t_1^1 \wedge x_0 \wedge x_1 \wedge x_2) \vee (t_1^0 \wedge (x_0 \vee x_1 \vee x_2)) \\
t_0^0 &= \neg t_0^1 \\
y_0 &= (t_1^1 \wedge t_0^0 \wedge x_1 \wedge x_2) \vee (t_1^0 \wedge t_0^1 \wedge (x_1 \vee x_2)) \vee (t_1^0 \wedge t_0^0) \\
y_1 &= (t_1^1 \wedge t_0^0 \wedge x_0 \wedge x_2) \vee (t_1^0 \wedge t_0^1 \wedge (x_0 \vee x_2)) \vee (t_1^0 \wedge t_0^0) \\
y_2 &= (t_1^1 \wedge t_0^0 \wedge x_0 \wedge x_1) \vee (t_1^0 \wedge t_0^1 \wedge (x_0 \vee x_1)) \vee (t_1^0 \wedge t_0^0)
\end{aligned}$$

Next to the negative side. Suppose that using m NOT gates we can invert n inputs. Consider the $n+1$ input vectors x^0 to x^n defined thus (here α runs from 0 to n):

$$x_\alpha^i = \begin{cases} 1 & \alpha < i \\ 0 & \alpha \geq i \end{cases}.$$

For example, if $n = 3$ then $x^0 = 000$, $x^1 = 100$, $x^2 = 110$ and $x^3 = 111$. For input x denote by $N(x)$ the m inputs to NOT gates, and by $O(x)$ the n outputs of the circuit. Since $N(x)$ can get only 2^m values, if $n+1 > 2^m$ there are $i < j$ with $N(x^i) = N(x^j) = y$ for some vector y. However, given that $N(x) = y$ the outputs $O(x)$ are monotone functions of the inputs. In particular, $1 = O(x^i)_i < O(x^j)_i = 0$, a contradiction. Thus $n \leq 2^m - 1$.

Finally, suppose that $n = 2^m - 1$. We will show that, in a sense, the solution we presented is the only one. In any solution, $O(x)$ must be monotone given $N(x) = y$ for any constant y we choose. In particular, in any chain in the cube $[0, 1]^n$ every vertex must have a different $N(x)$. We claim that this forces $N(x)$ to depend only on $\mathrm{HAM}(x)$.

To prove our claim, we use induction to prove the following property: if the cube $[0, 1]^n$ is colored with $n + 1$ colors in such a way that in each chain all colors are different, then the coloring depends only on the Hamming weight. This property is trivial for $n = 1$. Suppose next it holds for $n - 1$, and consider a coloring of $[0, 1]^n$ with $n + 1$ colors. Since $0^n \leq x$ for any x, $c(x) \neq c(0^n)$ for $x \neq 0^n$. Consider all vectors x satisfying $x_i = 0$ for some co-ordinate i. These form a copy C_i of $[0, 1]^{n-1}$ colored with $n - 1$ colors, and by induction this coloring must depend only on the Hamming weight. We claim that these n colorings are all compatible. Indeed, consider a maximal chain through δ_i passing at $\delta_i + \delta_j$ for some $i \neq j$ (here δ_i is the vector whose only non-zero co-ordinate is i). Replacing δ_i by δ_j, we see that the colorings of C_i and C_j must be compatible, completing the proof of the property.

We have shown that $N(x)$ depends only on $\mathrm{HAM}(x)$. Since $N(x)_0$ is monotone, we conclude that $N(x)_0$ must equal $\mathrm{HAM}(x)_{m-1}$. Given $N(x)_0 = h_0$, $N(x)_1$ must be monotone, and so $N(x)_1$ must equal $\mathrm{HAM}(x)_{m-2}$. Continuing this way we see that $N(x)$ is simply the reverse of $\mathrm{HAM}(x)$. In this sense the solution we presented is unique.