This is a small rephrasing of Step 1 of the proof of Theorem 14.4 in the text, showing that PARITY is not in \( \text{AC}_0^0(3) \).

Roughly speaking, Step 1 shows that the gates in an \( \text{AC}_0^0(3) \) circuit with \( n \) inputs can be well approximated by a polynomial of degree \( \sqrt{n} \) over the field \( \text{GF}[3] = \{0, 1, -1\} \). Then Step 2 shows that \( \oplus(x_1, \ldots, x_n) \) cannot be so approximated.

Here is the precise statement and proof of Step 1.

**Lemma 1** Let \( C \) be a \( \text{AC}_0^0(3) \) circuit of depth \( d \) with \( n \) inputs \( x_1, \ldots, x_n = \vec{x} \) and \( S \) gates. Then for all \( \ell \geq 1 \) there is a polynomial \( \tilde{g}(\vec{x}) \) over \( \text{GF}[3] \) of degree at most \( (2^{\ell})^d \) such that \( \tilde{g}(\vec{x}) = g(\vec{x}) \) for all but a fraction of at most \( S/2^\ell \) inputs \( \vec{x} \in \{0, 1\}^n \), where \( g \) is the output gate of the circuit.

**Proof:** We use induction on the depth \( d \), fixing \( \ell \geq 1 \). The induction hypothesis states that \( \tilde{g}(\vec{x}) \in \{0, 1\} \) for all \( \vec{x} \in \{0, 1\}^n \), and the approximating polynomial \( \tilde{g}(\vec{x}) \) has degree at most \( (2^\ell)^d \). We will show that for each gate \( g \), the approximating polynomial \( \tilde{g}(\vec{x}) \) introduces at most a fraction \( 1/2^\ell \) more errors on the input set \( \{0, 1\}^n \).

The base case is \( d = 0 \), where \( g = x_i \) for some input \( x_i \). Then define the polynomial \( \tilde{g}(\vec{x}) = x_i \). Then \( \tilde{g} \) has degree \( 1 = (2^\ell)^0 \), and there are no errors.

For the induction step, assume that gate \( g \) has depth \( d \geq 1 \), and the induction hypothesis applies to all gates \( f \) of depth at most \( d - 1 \).

**Case 1:** \( g = \neg f \) for some gate \( f \) in the circuit.
Then define \( \tilde{g} = 1 - \tilde{f} \). Then no new errors are introduced, and \( \tilde{g} \) and \( \tilde{f} \) have the same degree.

**Case 2:** \( g = \text{MOD}_3(f_1, \ldots, f_k) \). Then define

\[
\tilde{g} = \left( \sum_{i=0}^{k} \tilde{f}_i \right)^2
\]

The reason for squaring the sum is that \((-1)^2 = 1\) and \(0^2 = 0\), so that \( \tilde{g}(\vec{x}) \) is in \( \{0, 1\} \) for all \( \vec{x} \in \{0, 1\}^n \). Then no new errors are introduced, and the degree of \( \tilde{g} \) is double the largest degree of any \( \tilde{f}_i \). Thus the degree of \( \tilde{g} \) is at most \( 2 \cdot (2^\ell)^{d-1} \leq (2^\ell)^d \).

**Case 3:** \( g \) is the AND or OR of other gates. (This is the difficult case.) Without loss of generality we may assume \( g \) is an OR gate, since the AND gates can be expressed as OR gates using DeMorgan’s laws. Thus

\[
g = \bigvee_{i=1}^{k} f_i
\]

We could try setting \( \tilde{g} = 1 - \prod_{i=1}^{k} (1 - \tilde{f}_i) \). This would introduce no new errors, but the degree would increase by a factor of \( k \), which could be much more than the factor of \( 2^\ell \) which we allow.

So instead we make an approximation, and introduce errors.
CLAIM: If there is at least one \( j \) such that \( \tilde{f}_j \neq 0 \), then given a random \( k \)-tuple \( T = (c_1, \ldots, c_k) \) of field elements in \( \text{GF}[3] \), the probability that \( \sum_i c_i \tilde{f}_i = 0 \) is at most 1/3.

Proof: The set \( \{(c_1, \ldots, c_k) \mid \sum_i c_i \tilde{f}_i = 0\} \) is closed under linear combinations, and hence is a subspace of the vector space of \( k \)-tuples of elements in \( \text{GF}[3] \). It is a proper subspace, since \( \tilde{f}_j \neq 0 \). Every vector space of dimension \( m \) over \( \text{GF}[3] \) has \( 3^m \) elements. Hence this subspace (consisting of the ‘bad tuples’) consists of at most \( 3^{k-1} \) of the \( 3^k \) tuples \( (c_1, \ldots, c_k) \).

This proves the CLAIM. Now randomly pick \( \ell \) tuples \( T_1, \ldots, T_\ell \) with \( T_i = (c_{i1}, \ldots, c_{ik}) \) (with each \( c_{ij} \in \text{GF}[3] \)) and define

\[
p_i(x) = \left( \sum_j c_{ij} \tilde{f}_j \right)^2
\]

Now define

\[
\tilde{g}(x) = 1 - \prod_{i=1}^{\ell} (1 - p_i)
\]

Since \( p_i(x) \in \{0, 1\} \) for \( x \in \{0, 1\}^n \) it follows that \( \tilde{g}(x) = \text{OR}(p_1, \ldots, p_\ell) \). Then the degree of \( \tilde{g} \) is at most \( 2\ell \cdot (2\ell)^{d-1} = (2\ell)^d \). Also, by the CLAIM, for each \( x \in \{0, 1\}^n \)

\[
\Pr_x[\tilde{g}(x) \neq \bigvee_{i=1}^k \tilde{f}_i] \leq 1/3^\ell < 1/2^\ell
\]  \hfill (1)

Let’s say a pair \( (x, (T_1, \ldots, T_\ell)) \) is BAD if \( \tilde{g}(x) \neq \bigvee_{i=1}^k \tilde{f}_i \). By (1) at most a fraction \( 1/2^\ell \) of the pairs are BAD, and hence there exists a choice \( (T_1, \ldots, T_\ell) \) which is not BAD for all but a fraction \( 1/2^\ell \) of elements \( x \in \{0, 1\}^n \). We use this choice for \( T_1, \ldots, T_\ell \), so \( \tilde{g} \) differs from \( \bigvee_{i=0}^k \tilde{f}_i \) on at most a fraction \( 1/2^\ell \) of inputs \( x \).

Finally note that there are \( S \) gates in the circuit, and for each gate \( g \), the approximating polynomial \( \tilde{g} \) introduces at most a fraction \( 1/2^\ell \) further errors on the input set \( \{0, 1\}^n \). Hence the output approximating polynomial \( \tilde{g} \) differs from the output gate \( g \) on at most a fraction \( S/2^\ell \) of the inputs.

Using the Lemma, we choose \( \ell = (1/2)n^{1/2d} \). Then the degree of \( \tilde{g} \) is at most \( (2\ell)^d = (n^{1/2d})^d = \sqrt{n} \). Step 2 shows that no polynomial over \( \text{GF}[3] \) of degree \( \sqrt{n} \) agrees with \( \oplus(x_1, \ldots, x_n) \) on more than a fraction of 49/50 of inputs \( x \in \{0, 1\}^n \), so the fraction of errors is at least 1/50. Thus if the circuit has depth \( d \) and \( S \) gates and computes \( \oplus(x) \), then

\[
S/2^\ell \geq \text{error fraction of } \tilde{g} \geq 1/50
\]

so \( S \geq (1/50)2^{(1/2)n^{1/2d}} \).