VLSI DESIGN
1996, Vol. 5, No. 1, pp. 89–100
Reprints available directly from the publisher
Photocopying permitted by license only

Zener Zap Anti-Fuse Trim in VLSI Circuits

DONALD T. COMER*

Brigham Young University, Provo, UT 84602

(Received 23 October 1994; In final form 10 April 1995)

This paper presents an overview of *Zener zap* anti-fuse trim as used to achieve improved accuracy in precision integrated circuits. Because this technology spans design and manufacturing, elements of design, layout, processing, and testing are included. The mechanism is defined and typical applications are discussed. Layout considerations of anti-fuse devices are summarized and complex trim networks and multiplexed control methods are presented. Both bipolar and CMOS process implementations are considered. The paper also contains a bibliography which includes U.S. patents, which make up a large part of the technical documentation of this technology.

Keywords: Zener zap, anti-fuse, trim, precision integrated circuits, mixed signal circuits, VLSI circuits

1. INTRODUCTION

The accuracy of analog integrated circuits is typically limited by poor control of the absolute value and matching tolerances of the integrated devices created in the fabrication process. Although ratiometric design principles are used where possible, mismatching in ratios can cause a loss of accuracy and poor yield to high accuracy specifications.

Over the years, "trim" techniques have been developed to improve the accuracy and yield of integrated circuits. In this context trimming refers to making adjustments to the integrated circuit after its fabrication has been completed. Trimming is typically done at the wafer level on individual die but may even be performed on die after packaging. Trimming is typically targeted to achieve improvement in key parameters such as the offset voltage of an op-amp, the absolute value of a reference voltage, or a specific delay time in a cascade of logic gates.

The advent of mixed-signal VLSI has created a renewed interest in inexpensive methods of trimming at the wafer level to improve yield. This is because the yields of digital circuits typically exceed those of analog circuits constructed on the same process [13]. Since most mixed-signal integrated circuits are size-dominated by digital circuits, the failure of a small analog section of a chip to meet accuracy requirements can have drastic implications on the yield economics of a mixed-signal design.

A widely used method of trimming integrated circuits is laser trimming of thin film resistors [1]. By

*Fax: (801) 378-6586 E-mail: dtc3@ee.byu.edu.

selectively evaporating a small portion of the film resistor with a laser beam, its effective value can be increased. This method is used, for example, in achieving high linearity in D/A converters, where several resistors must be trimmed to prescribed ratios [2].

An alternative to laser trimming to achieve precision in integrated circuits is the anti-fuse approach which is the subject of this paper. This method is applicable to both bipolar and CMOS designs and is especially suited to mixed-signal ASIC products which do not lend themselves to laser trimming.

The term *anti-fuse* is used to describe an element which initially appears as an open circuit but can be made to approach a short circuit by forcing conduction of a high current for a short duration. Anti-fuse devices may be created in integrated circuits by three distinct methods. The first method utilizes a dielectric material to insulate two conductors, which can be effectively shorted together by electrically breaking down the dielectric [14–15]. A second method is based upon the use of amorphous silicon material, which in its normal state is an adequate insulator but can be rendered conductive by the application of sufficiently strong fields and current flow [17–18]. The third method creates anti-fuse devices by forcing a temporary avalanche breakdown in a *P-N* junction sufficient to cause localized heating and subsequent migration of metal across the junction. This method of creating an anti-fuse came to be known in industry jargon as "Zener zap" and in fact the Zener zap terminology is now used in technical literature as well [9–11].

Although it is beyond the scope of this paper to make a quantitative comparison of the three anti-fuse approaches, it can be noted that the dielectric anti-fuse is dependent upon utilizing thinner oxides than are available in most standard bipolar integrated circuit processes and therefore tends to be used primarily in CMOS and BiCMOS fabrication processes. The amorphous silicon approach utilizes a material that is not commonly used in either bipolar or MOS integrated circuit processes and its use requires a special processing step in the fabrication cycle. The Zener zap method requires a *P-N* junction device with a low to moderate junction breakdown voltage. Historically the Zener zap method has been used primarily on bipolar processes, where its use requires no additional or special processing steps. The primary focus of this paper is on the Zener zap method, although the basic resistor link trim methods discussed could be used with other anti-fuse technologies.

2. DESCRIPTION OF THE ZENER ZAP MECHANISM

Fig. 1 shows a top view and a cross-section of a device used in a typical bipolar process to implement the Zener zap anti-fuse. In this case the cathode of the diode is created as a circular diffusion with a circular contact. The cathode is heavily doped *N*-material (typically 1–3 microns deep and 5–10 ohms per square) and is installed at the same process step as the emitter of a complete NPN device. The anode of the Zener device is created by a moderately doped *P*-diffusion (typically 3–5 microns deep and 100–200 ohms per square) and is installed at the same process step as the base of a bipolar NPN device.

The equation governing the breakdown voltage of the *P-N* junction is given by

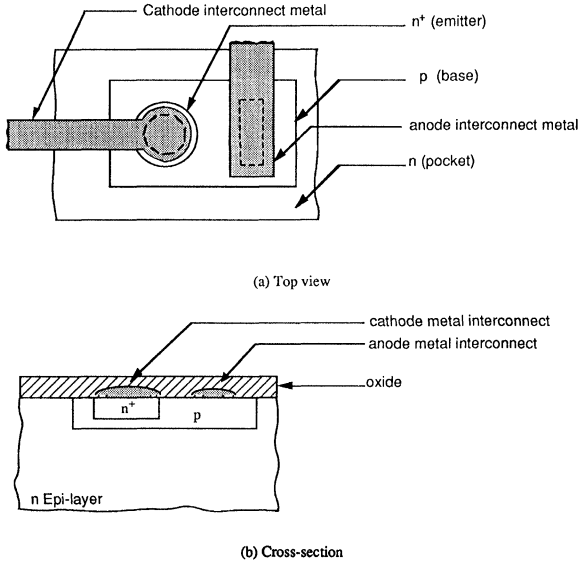


FIGURE 1 Typical Bipolar Zener Device

$$BV = \frac{\epsilon(N_A + N_D)\varepsilon_{\text{crit}}^2}{2qN_AN_D}, \quad (1)$$

where ϵ is the permittivity of silicon, N_A and N_D are the acceptor and donor doping densities of the *P-* and *N-* materials respectively, q is the charge of an electron, and ε_{crit} is the maximum field that can be impressed across the depletion region of the *P-N* junction without leading to avalanche breakdown [8].

For the case of a diode doped as suggested, $N_D >> N_A$, (1) may be simplified to

$$BV \approx \frac{\epsilon\varepsilon_{crit}^2}{2qN_A}. \quad (2)$$

Equation (2) indicates that the *P-* material doping will determine the junction breakdown and that a higher doping will result in a lower breakdown voltage. The doping levels used in most bipolar and Bi-MOS processes create a breakdown voltage in the range of 6 to 7 volts. Furthermore, because most fabrication processes create a doping profile that results in maximum density at the surface of the doped material, it follows that the voltage breakdown will occur at the surface of the vertical *P-N* junction.

The power dissipated by the junction at breakdown is given by

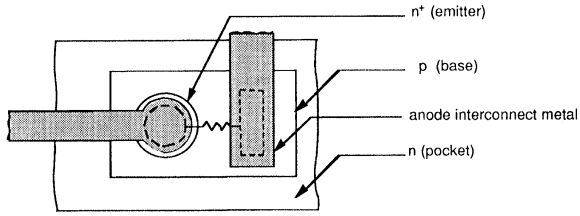
$$P = BV \cdot I, \quad (3)$$

where I is the current conducted across the junction during breakdown. If no limit is placed upon I, the junction will heat very rapidly and can be destroyed by a number of mechanisms. On the other hand, if I is limited, the power P will cause localized heating around the area where the current is concentrated. If I is applied for a fixed time, sufficient heating can occur to cause migration of atoms of the metal interconnect from the cathode terminal to the anode terminal of the diode along the path of breakdown current. This migration manifests itself as a trace of metal (typically aluminum) imbedded in the silicon along the path of I (near the surface).

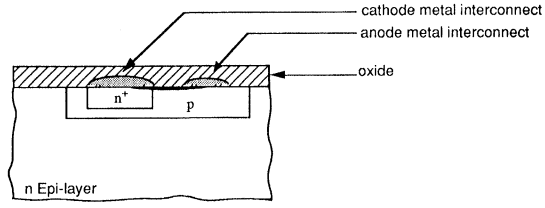
The mechanism by which the transport of metal atoms takes place by momentum exchange with conducting electrons is investigated in the literature as "electromigration" and is a very important factor in the reliability analysis of integrated circuit metalization [9]. It occurs in metal lines at high current densities and elevated temperatures and consists of the movement of metal atoms toward the positive terminal of the conductor. Electromigration of aluminum generally occurs at current densities in excess of 105 A;shcm^2 and at temperatures in excess of 100°C.

Fig. 2 shows a top view and a cross section of a Zener zap device which has been fused by metal migration. It is worth noting that the fused metal stripe occurs on and slightly below the surface of the silicon but is below the layer of passivation normally used to seal out contaminants. No damage is caused to the passivation and the result is a structure with no diminished reliability as a result of the fusing mechanism.

The amount of current required to cause fusing will vary with the process and device size. Small devices constructed with no special optimizations for this application on a typical bipolar process will require zap currents in the neighborhood of 100–200 mA for a duration of a few milliseconds. It is possible to design structures for a given process that will fuse



(a) Top view



(b) Cross-section

FIGURE 2 Zener Device After Fusing

at lower currents, as discussed in the following paragraphs. However, in all cases it is necessary to limit the current and control the maximum heating of the junction in order to create a reliable fused junction.

The fusing phenomenon as it is observed to occur in most processes may be divided into two phases. The first is the initial breakdown and heating phase in which the metal interconnect atoms are mobilized and begin to flow across the junction. The second phase involves the carrying of a sufficient number of metal atoms to create a low resistance path through the silicon. Although both phases can be combined, it has been observed that less damage to the junction, passivation and surrounding metal will be caused by initiating the first phase with a high peak current of short duration, followed by a lower current of longer duration. Fig. 3 shows idealized current wave-forms to implement the mechanism in two phases.

The resistance of a fused link on a typical process using Zener diodes and aluminum interconnect metal is in the neighborhood of 10 ohms.

3. TRIM APPLICATIONS

In the most elementary application of resistor trimming by Zener zap, the Zener element, Z_1, is placed in parallel with a resistor, ΔR, as shown in Figure 4a. Both terminals of the Zener element are accessible by

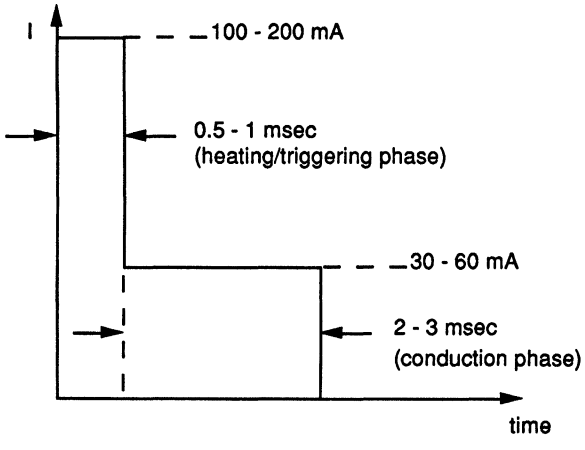


FIGURE 3 Idealized 2-phase Implementation of Zener zap

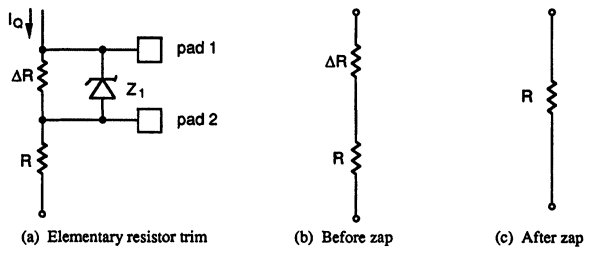


FIGURE 4 Single Link Resistor Trim

bonding pads of the integrated circuit. In cases where the trim increment is intended to be a small percentage of the total resistance, the resistance ΔR is made small compared to the total resistance of the string ($\Delta R << R$).

It is necessary that the Zener element be placed in the circuit such that it does not conduct current in the normal operating mode. For example, the current flow I_Q, existing in the resistive branch to be trimmed, should be in such a direction to reverse-bias the element Z_1 and not to exceed the Zener breakdown. That is,

$$I_Q \Delta R < BV. \tag{4}$$

Under this condition, the Zener element will not affect normal current flow in the branch and may be considered an open circuit (Fig. 4b) as far as its effect on normal biasing. However, if avalanche migration is induced in the Zener element, then the terminal resistance is typically around 10 ohms and is usually sufficiently small to be considered a short circuit. Figure 4c shows the equivalent circuit after the zap of Z_1.

In applications where a wider range of trim is required, it is possible to cascade trim elements and resistors as shown in Fig. 5. This allows a sharing of the bond pads of the Zener elements. For N trims in a resistor string, $N + 1$ bond pads will be required.

If individual pads are provided as shown in Fig. 5 to independently access any of the trim elements then all possible combinations of trim may be achieved. For N trim elements there are $2^N - 1$ trim combinations. It follows that if the trims are binary weighted (say with ΔR_1 being the lowest-order trim) all integer combinations of the lowest-order trim may be

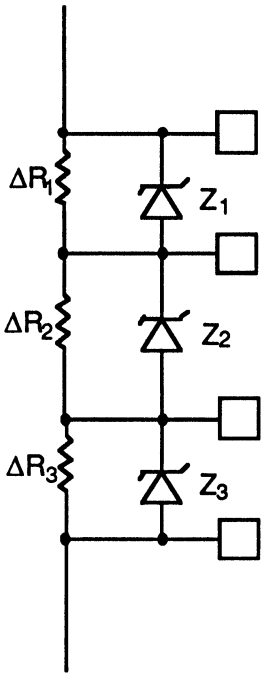


FIGURE 5 Cascaded Trim Elements

achieved up to $(2^N - 1)\ \Delta R_1$. The trim range, R_T, may then be determined as the maximum resistance change when all resistors are shorted:

$$R_T = \Delta R_1(2^N - 1). \tag{5}$$

The above results can be summarized as follows: Given a trim string as shown in Fig. 5 where the lowest order trim is ΔR_1 and $\Delta R_2 = 2\Delta R_1$, $\Delta R_3 = 2\Delta R_2$, etc., then the resolution of the trim is ΔR_1, and the range of trim is $\Delta R_1\ (2^N - 1)$. The untrimmed error, ε_U, of a resistor string may be defined as the difference between the untrimmed resistance and the desired value of the resistor string. The error remaining after trim, ε_T, may be expressed

$$\varepsilon_T = \varepsilon_U - \Delta R_1 M, \tag{6}$$

where M is one of the $2^N - 1$ trim combinations, chosen on the basis of producing the smallest value of ε_T.

The value of error remaining after trim will depend upon the initial error and will range between zero and a value $\varepsilon_T(max)$ given by

$$\varepsilon_T(max) = \frac{\Delta R_1}{2}. \tag{7}$$

The resistor trimming scheme illustrated in Fig. 5 provides *unilateral* trimming; that is, it is only possible to *decrease* the value of the resistor string but not to increase it. In applications where a resistor is to be trimmed to fixed precision, starting with an allowed variation that may be positive or negative, it is necessary to skew the untrimmed value of the resistor towards the larger extreme. For example, if we assume that the untrimmed resistor R_U has a nominal value, R_{nom}, and a tolerance, $x\%$, where

$$R_U = R_{nom} \pm x\%, \tag{8}$$

then a trim string using the method suggested by Fig. 5 would require an R_U value of $R_{nom} + x\%$ and a range of $-2x\%$. The number of trims could be determined once the desired resolution was known.

In certain applications, bilateral trimming may be achieved by virtue of the polarities of different trim strings. For example, the trimming of op-amp input offset voltages may be accomplished by the scheme shown in Fig. 6, where two resistor trim strings are

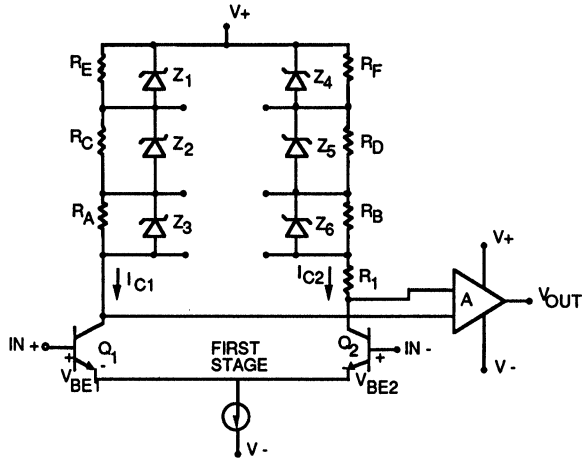


FIGURE 6 Bilateral Trim Application

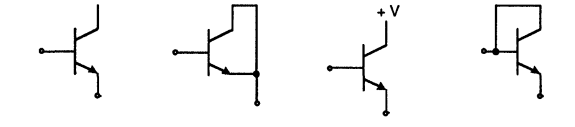
used as load resistors of a differential amplifier. In this case, each resistor trim string produces an opposite polarity change in offset and the initial design may be targeted to be perfectly balanced in the untrimmed state.

4. DEVICE LAYOUT AND DESIGN CONSIDERATIONS

In a conventional bipolar process it is a common practice to use the emitter-base junction of a standard NPN transistor device as a Zener zap fuse element. In other cases, especially where the zap trim is an integral part of the design, devices intended specifically for use as Zener zap fuses are custom designed. However, the layout of these devices must be consistent with the layout design rules for the process being used. In general, a process that allows very small device features will provide a good fuse element. Typically, device sizes are governed by the operating voltage requirement of the design. A general purpose op-amp process with an operating voltage requirement of 30 volts will require larger devices than a process intended to fabricate 5 volt logic circuits. However, within a set of design rules for a given process it is possible to optimize the structure for use as a fuse element.

An optimal fuse element is defined as one which can be fused most cleanly and reliably at the lowest value of current and power, without damage to the surrounding passivation or metal. The following list can be used as a guide to layout-design an optimal fuse element for a given process:

1. Metal traces feeding the fuse elements should be made sufficiently wide so that they will not be damaged anywhere along the path of high current conduction due to electromigration effects within the metal itself. Most processes use rules to guard against electromigration over a long period of time by establishing a current density limit for metal traces. For example, a typical rule for each 10,000 Å thick aluminum metal would be 1 micron of width for each milliampere of DC current to be carried. It is unrealistic to use this type of rule for the conduction of zap fuse currents which occur as single incidents for durations of only milliseconds. As a rule of thumb, currents may be increased by a factor of 20–30 over DC rules for zap fuse currents.
2. Once a minimum metal width is established for zap fuse devices this width should be used as a maximum when feeding traces into the terminals of the device. That is, extra wide busses such as ground or supply busses should not be connected directly to the terminals of the zap fuse devices. This is to avoid any local "heat-sinking" effects of the bus during the fusing process.
3. Devices should be biased so that no current flows in the substrate during zap fusing. This may be accomplished by letting the substrate float or by ensuring that it is biased more negatively than the pocket of the element being fused.
4. If an NPN bipolar transistor device is employed as the fuse element, options are available as to how to connect the collector (pocket) terminal. If the zap element is placed in the circuit at a location where its trimmed link sees primarily a DC bias, it is acceptable to allow the collector to float as shown in Fig. 7a. If there is concern about DC signals being capacitively coupled to the pocket and causing parasitic pocket-to-substrate currents, then the collector should be tied to a fixed bias potential. It is usually convenient to tie the collector to the emitter of the device as shown in Fig. 7b. In other cases, it is acceptable to tie the collector to the positive supply (Fig. 7c). The conventional collector-base diode connection (Fig. 7d) should be avoided as this will substantially increase the fuse current.



(a) Floating collector (b) Collector-emitter connection (c) Collector to positive supply (d) Conventional diode connection (to be avoided)

FIGURE 7 Transistor / Zener Connections

5. A round emitter is generally preferable to a square or rectangular one. If a round emitter is used, the zap channel will consistently occur at the tangent point on the circle closest to the base contact as shown in Fig. 8a. If a rectangular emitter (Fig. 8b) is used, the zap channel will typically occur at one corner or the other of the device emitter but may occur anywhere along the emitter edge nearest the base contact. Aside from producing a zap channel with a consistent location, the round emitter device typically will fuse at a slightly lower current. Also, if a judicious waiver of metal spacing is to be used for the zap devices (for example, allowing base-emitter metal to be 1–2 microns closer than normal metal-to-metal minimums), the closer spacing only need be implemented by a small "metal flap" of extra emitter metal as shown in Fig. 9. This typically will not cause any yield loss due to metal bridging and will decrease fuse current as much as 10%. In general, the closer the spacing between emitter and base metal, the lower the fuse current will be.
6. Any device structure that will maximize current density in the zap channel will reduce the fuse current. An example of a structure that can be built specifically for a fuse element is shown in Fig. 10. In this case, no attempt is made to place the emitter $N+$ diffusion inside the $P-$ diffusion as in a conventional transistor device. Instead the $N+$ diffusion is primarily setting in the pocket. This causes an electrical connection between the cathode and the pocket (similar to shorting the emitter and collector of an NPN device together). A small overlap is provided between the circular $N+$ diffusion and the arrow-shaped $P-$ diffusion. The zap channel is then created within the overlap of the $P-$ and $N+$ diffusions. A structure of this type may reduce the fuse current by a factor of 20–30% for a given process and layout design rules.
7. Zener zap trim devices may be constructed on either bipolar or CMOS fabrication processes. The device structure suggested in Fig.10 may be implemented on a standard P-well CMOS process, where the anode is created using the P-source/drain implant, and the cathode is created using the N-source/drain implant. Both electrodes are placed within a P-well to achieve isolation from the substrate. Such an element has, in fact, been constructed on the MOSIS 2-micron P-well process and was found to fuse at a current of approximately 50mA [12].

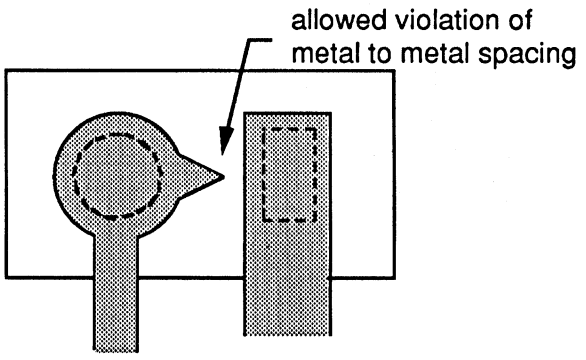
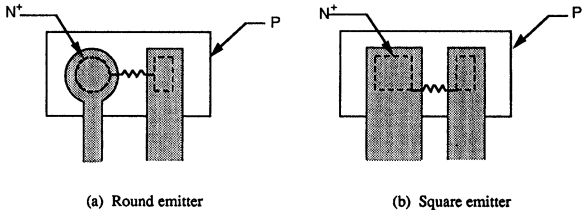


FIGURE 9 Emitter Metalization for Lower Zap Current



(a) Round emitter (b) Square emitter

FIGURE 8 Typical Zap Channels

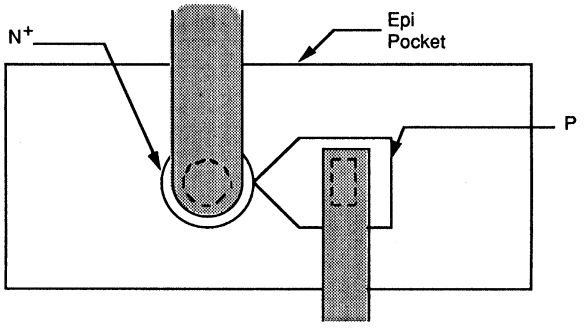


FIGURE 10 Custom Designed Zener zap Device

5. DESIGN OF RESISTOR STRINGS

A resistor string to be link trimmed over a wide range and with a small resolution is normally constructed from binary-scaled resistors starting from the smallest resistor, which provides the desired resolution, and including sufficient links to make up the desired range. However, extremely low or high resistor values can create problems by interfering with the zapping mechanism. For example, assume that a trimmable resistor string is to be designed with a unit resolution of 125 Ω and a trim range of 63 units or 7875 Ω. Following the binary weighting of resistor values suggested earlier would result in the resistor string shown in Fig. 11. This design would normally be unacceptable because of the diversion of zap current by the low value resistors.

Since the elements of the resistor string are shunted across the fuse elements, the resistor metal interconnect lines will conduct current at the time a zap signal is applied to a given element. For low value resistors, this current can be excessive and can cause damage to either the resistor or its metalization path. For example, if the zap voltages are allowed to go as high as 30 volts, the current in R_1 of Fig. 11 could rise to a value of 240 mA during the zap of Z_1. Not only might this level of current cause damage to R_1, but it may put an excessive load on the driver of the probe-interface. In general, it is wise to limit the shunt resistor current to about 20% of the required zap current. For a typical zap current of 100 mA and a driver swing of 30 volts, the maximum shunt resistor is determined to be 30 volts/20 mA = 1.5kΩ. Thus, all resistors except R_6 and R_5 in the string of Fig. 11 would be too small.

Fig. 12 shows an alternative resistor string to that of Fig. 11 which provides the same trim range and resolution, but adds additional resistors to limit the currents during zap. The alternative approach shown

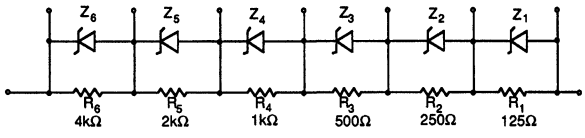


FIGURE 11 Example of Improper Trim String

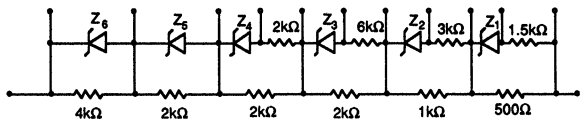


FIGURE 12 Modified Resistor String

in Fig. 12 may be thought of as a type of resistor ladder. The calculation of resistor elements is intuitive and is not detailed here. However, it is a simple matter to verify the equivalence of the trim range and resolution of the resistor strings of Figs. 11 and 12. It should be pointed out that the two resistor strings are not equivalent from a total resistance point of view. The untrimmed value of the string of Fig. 11 is 7875 Ω while the comparable value of the string if Fig. 12 is 11.5 kΩ. This shift in range normally can be accounted for in the overall circuit design. In general, the concept of changing a link value by paralleling it with a larger resistor provides for improved resolution and generally may be used in any case where small resistor trim increments are required.

It is also important to note that adding additional resistors in series with the fuse elements increases the number of access nodes (typically bonding pads) required to trim the resistor string. This is because the access nodes cannot be shared between adjacent fuse elements as in Fig. 11. For example, 7 access nodes are required to trim the resistor string of Fig. 11 while 11 access nodes are required to trim the modified resistor string of Fig. 12. If the extra resistors in the modified string are rearranged as shown in Fig. 13 to allow some of the zap elements to be commoned, the number of access nodes can be reduced from 11 to 9 as shown.

A second design caution for trim strings that involve large resistors is to ensure that none of the Zener elements can become conductive during normal circuit operation. If the quiescent current I_Q in the resistor string is unidirectional then the zener elements should be oriented so that they are normally in

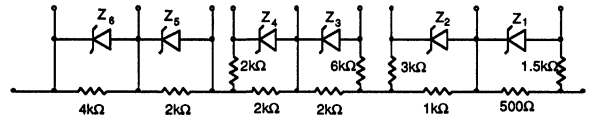


FIGURE 13 Alternative Modified Resistor String

a reverse-biased condition with the current I_Q limited as stated in (4). If other constraints require placing the Zener elements in the opposite polarity relative to I_Q, then the bias constraint requires that $I_Q \Delta R$ be less than the diode turn-on threshold at the highest expected operating temperature.

In early analog products which employed Zener zap, access to the anti-fuse links was provided through the bonding pads of the integrated circuit chip. Control circuitry was then interfaced through probes contacting the bond pads at wafer sort. This remains the simplest method of trim control for adjustments requiring only a few trims. However, a major limitation in the number of link trims that are permitted for a given product is the number of bond pads available. Because of this limitation, methods of reducing the number of bonding pads required for trim access have been developed. Some of the prominent multiplexed methods are described in the following section.

6. MULTIPLEXED CONTROL METHODS

The primary goal of multiplexed methods of control for Zener zap is to achieve more area efficient use of bond pads. Also, since the bond pad method limits access to trimming at wafer probe, it is not possible to perform "after package" trim which is very desirable for some products. Thus, a secondary goal of multiplexed control is to allow the possibility of after-package trim. Figure 14 shows a very simple method of multiplexing a single bond pad to control two Zener zap anti-fuse elements. In this case, selection is based upon polarity of the zap signal [7].

A positive voltage applied at the control pad will force element Z_2 into avalanche breakdown with current being supplied by element Z_1, which behaves as a forward-biased diode in this case. A negative voltage applied to the control pad will force Z_1 into avalanche breakdown with current being supplied by Z_2. Either element can be independently fused with no limitation placed upon sequence. For example, if Z_1 is fused initially by a negative pulse, Z_2 can then be

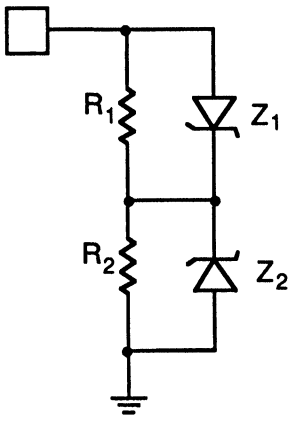


FIGURE 14 Polarity Multiplexed Control Circuit

fused by a positive pulse with current being supplied through the fused Z_1. This method of multiplexing can potentially reduce the number of bond pads by a factor of 2, but is limited to trims with a small range since its application is complicated when used with the large resistor ladders as discussed earlier.

Figure 15 shows a slightly more complex method of multiplexing a single bond pad to control three fuse elements based upon a combination of polarity and sequence of zap signals [11]. The conduction of a high current from pad 1 to ground will fuse Z_3 and cause resistor R_3 to shunt R_4. During fusing of Z_3, transistor Z_2 will conduct current through its collector and effectively bypass element Z_1. Therefore, element Z_1 will remain intact following the first zap current pulse applied to Z_3. Also, element Z_2 will be

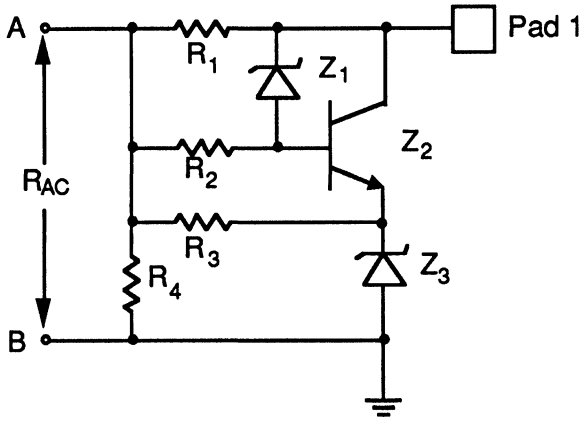


FIGURE 15 Polarity - sequence Multiplexed Control Circui

undamaged since current flow through the device will be in its normal collector-base-emitter (forward-biased) direction.

After Z_3 has been fused, there is a low resistance path for current to flow in the reverse direction. A second high current pulse flowing from ground to pad 1 will cause fusing of Z_2, causing R_2 to shunt R_4.

Finally, a third high current pulse applied so that current flows from pad 1 to ground will fuse Z_1, causing resistor R_1 to be shunted across R_4. In this case, the control current will flow through already shorted elements Z_2 and Z_3. There will be no collector current flowing in Z_2 because transistor action will not take place after Z_2 has been zapped.

The circuit of Fig. 15 illustrates the use of the collector terminal of a Zener zap element connected as a transistor to selectively "protect" other zap elements for certain polarities and sequence of zap pulses. By proper selection of polarity and sequence of current pulses, three fuse elements can be accessed from a single pad. In order to achieve simultaneously the most range and resolution for a given number of independent elements, it is common practice to binary weight each zap element as discussed earlier. However, in the multiplexed control circuit of Fig. 15, since access to a given trim is dependent upon a prior trim, the weighting of the trim increments must be changed in order to prevent any inaccessible combinations. A single trim pad using the circuit of Fig. 15 must therefore be designed for 3 equal trims of T_1, providing a trim range of $3T_1$ and a resolution of T_1.

It is possible, of course, to replicate multi-zap circuits in order to increase the overall trim range. To do so, the multi-zap circuits may be connected in a cascade or in parallel. The range is then increased more than proportionally to the number of the circuits employed. For example, a single multi-zap circuit will provide a three to one advantage in trim range over that obtainable with a single trim element. However, if three independent multi-zap circuits are cascaded, the first can be weighted T_1 with a range of $3T_1$, the second at $4T_1$ (to achieve the next desired trim increment) with a range 12T1 and the third at $16T_1$ with a range of $48T_1$. The total range resulting from 3 pads is then $63T_1$ compared to a range of $7T_1$ for three independent zaps using a straight binary, non-multiplexed method.

The trim control circuitry described above may be rendered in integrated circuit form by conventionally diffusing three discrete NPN transistors into a substrate, and interconnecting the three to form the circuit of Fig. 15. Area savings may be obtained, however, by using a unique layout of the active elements as shown in Fig. 16.

The special geometry is intended to provide the same function as would individual fabrication of the three devices Z_1, Z_2, and Z_3 of Fig. 15, but is integrated within a smaller area on a silicon chip. Figure 16a depicts the metal interconnect while Fig. 16b depicts the isolation, base and emitter diffusions necessary to make the device. The figures are not necessar-

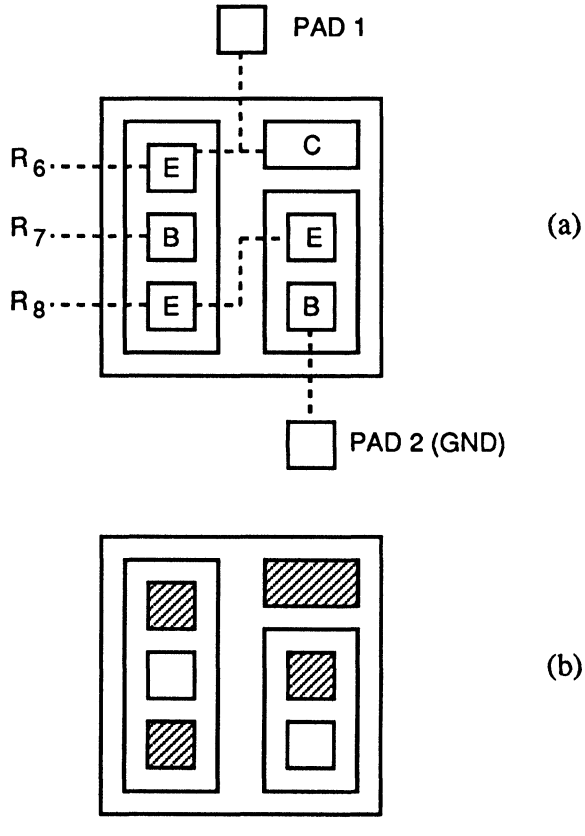


FIGURE 16 Cell Layout of Multi-zap Circuit

ily to scale, but show only the relative positioning of the devices which makes possible the desired reduction of area on the chip.

7. LOGIC MATRIX CONTROL

The application of logic encoding/decoding schemes to the control of fusible link trimming is complicated by the fact that the output of the control device must actually deliver the fusing current. It is not sufficient to merely identify the desired node by logic decoding since a relatively high amount of power must be delivered to the node in order to effect fusing. One scheme which has been applied to trim control is the "matrix driver" approach described in [4–6]. In this case, a matrix of upper and lower drivers are constructed using NPN power devices as suggested in Fig. 17.

In this case, four Zener zap elements are controlled by a 2 × 2 matrix of drivers consisting of Q_1 and Q_2 as the "upper drivers"and Q_3 and Q_4 as the "lower drivers". A high voltage on one upper and one lower driver base will then select one of the four zap elements.

Although there is no gain in efficiency of control for a 2 × 2 matrix as shown since four access points are required to control four elements, higher order matrices can show an improvement in control efficiency. For example, a 6 × 6 matrix of drivers can access 36 zap elements with 12 access control nodes. In general, the number of access nodes required for an $N \times N$ matrix is $2N$ whereas the number of elements controlled is N^2. A control reduction factor F_C can be defined as the ratio of elements controlled to the number of control access nodes. For the matrix driver scheme just described

$$F_C = \frac{N}{2}. \tag{9}$$

Thus, an 8 × 8 matrix would show a control reduction factor of 4. However, it should be pointed out that the additional drivers required to implement the matrix scheme will typically require additional chip area and will decrease the effective control reduction factor. If we make the assumption that a driver element in the matrix decoding scheme requires roughly the same area as a pad and bond pads are assumed to be the primary access control nodes, then a matrix driver will require $2N$ drivers plus $2N$ bond pads, an effective area equivalent to $4N$ access nodes. Thus, the effective control reduction factor, F_c', in terms of decreased chip area for this case would become

$$F_c' = \frac{N}{4}. \tag{10}$$

Equation (10) indicates that only relatively large N matrices will show savings in area. For example, a 4 × 4 matrix would break even area wise, but would require additional circuit complexity.

An advantage of the matrix driver approach is that it allows a method of "after package trim". For example, the first integrated circuit product to use such an approach was a 12-bit D/A converter [4], where a 6 × 6 matrix was employed to control 36 trims. By utilizing a "threshold zener" in the base of the drivers, a type of level-sensitive multiplexing was created where the 12-bit DAC input lines could be operated at a higher than normal voltage to activate the matrix drivers. This was accomplished at the package level

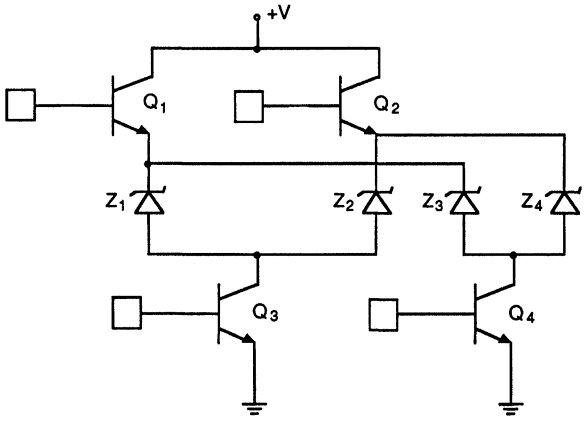


FIGURE 17 Matrix Driver Control Circuit

and thus the device could be trimmed after package, sealing, and burn-in.

8. SUMMARY

Anti-fuse technology is widely used in programming digital logic arrays but may also be used to perform trimming in precision analog circuits. It is especially important in mixed-signal applications as a means of improving yields of analog sections of a chip to levels that are comparable to the digital sections.

The three primary anti-fuse methods in use today are the dielectric, the amorphous silicon, and the Zener zap anti-fuse. Although Zener zap has historically been the technology of choice for the trimming of precision analog circuits, it is possible that other anti-fuse approaches may also be used. The advantage of the Zener zap method is that it may be used with most bipolar, CMOS and BiCMOS processes without requiring any special process steps. Its disadvantage is that it is not as area efficient as either of the other methods. Its area efficiency may be improved by the use of specially designed devices and the use of multiplexed control methods as discussed in this paper.

References

[1] M. I. Cohen, B. A. Unger, and J. F. Milkosky, "Laser Machining of Thin Films and Integrated Circuits," *Bell Syst. Tech. J.* 47, 385–405 (1968).

[2] S. Harris and D. Wagner, "Laser Trimming on the Chip," *Electronic Packaging and Production*, pp. 50–56, February, 1975.

[3] G. Erdi, "A Precision Trim Technique for Monolithic Analog Circuits," *IEEE Journal of Solid State Circuits*, Vol. SC-10 No. 6, pp. 412–416, December 1975.

[4] D.T. Comer, "A Monolithic 12-bit DAC," *IEEE Transactions Circuits and Systems*, CAS-25, 504–509, July 1978.

[5] D.T. Comer, et. al.: "Selectable Trimming Circuit for Use with a Digital to Analog Converter," U.S. Patent No. 4, 138, 681, Feb. 6, 1979.

[6] D.T. Comer, "Trimming Control Circuit for a Digital to Analog Converter," U.S. Patent 4, 131,884, February 14, 1977.

[7] R.C. Dobkin, "Integrated Circuit on Chip Trimming," U.S. Patent No. 4,225,878, September 30, 1980.

[8] Paul R. Gray, and Robert G. Meyer, "Analysis and Design of Integrated Circuits," John Wiley and Sons, 1984, p. 389.

[9] Alan Grebene, "Bipolar and MOS Analog Integrated circuit Design," John Wiley and Sons, 1984, Chapter 3.

[10] G. Erdi, "Integrated Circuit Having Permanent Adjustment Circuitry which Requires Low Adjustment Current," U.S. Patent No. 4,775,884, October 4, 1988.

[11] D. T. Comer, "Apparatus for Multiple Link Trimming in Precision Integrated Circuits," U.S. Patent No. 4,777,471, Oct. 11, 1988.

[12] D. T. Comer, "Fusible Link Trim of Precision Integrated Circuits," Research Report No. SET-VLSI-102.1, May 22, 1992. Penn State Harrisburg Middletown, PA.

[13] D. T. Comer, "Introduction to Mixed-Signal VLSI," Array Publishing, New York, ISBN 0-9638044-0-1, 1994.

[14] D.K.Y. Liu, K.L. Chen, H. Tigelaar, J. Paterson, S.O. Chen, "Scaled Dielectric Antifuse Structure for Field-Programmable Gate Array Applications," *IEEE Electron Device Letters*, Vol. 12, No. 4, pp. 151–153, April 1991.

[15] E. Hamdy, J. McCollum, S. Chen, S. Chiang, S. Eltoukhy, J. Chang, T. Speers, A. Mohsen, "Dielectric Based Antifuse for Logic and Memory ICs," *IEDM Tech. Dig.*, pp. 786–789, 1988.

[16] Jonathon Greene, "Antifuse Field Programmable Gate Arrays," *Proceeding of the IEEE*, Vol. 81, No. 7, July 1993, pp. 1042–1056.

[17] QUICKLOGIC Corporation, "1994 Data Book," 2933 Bunker Hill Lane, Santa Clara, CA.

[18] R. J. Wong and K.E. Gordon, "Conducting Filament of the Programmed Metal Electrode Amorphous Silicon Antifuse," *IEDM Tech. Dig.*, December, 1993.

Authors' Biography

Donald T. Comer received B.S., M.S. and Ph.D. degrees in Electrical Engineering from San Jose State University, the University of California at Berkeley and the University of Santa Clara respectively. He combined industrial work with teaching while he was professor at San Jose State University during the 1960s and 1970s. He has worked for industrial firms such as Precision Monolithics, IBM Corporation and Analog Devices, and has served as a consultant to various firms including Fairchild and Raytheon. He holds several patents and has published numerous articles and textbooks related to solid state and integrated circuits. Currently, he holds the Endowed Chair of Engineering at Brigham Young University.