Vladislav A. Vashchenko
Andrei Shibkov

ESD Design for Analog Circuits

Springer

ESD Design for Analog Circuits

Vladislav A. Vashchenko · Andrei Shibkov

ESD Design for Analog Circuits

Springer

Vladislav A. Vashchenko
National Semiconductor
Semiconductor Drive M/S E-155
Santa Clara, CA 95052
USA
vladislavv@att.net

Andrei Shibkov
Angstrom Design Automation
1336 Fruitdale Ave. A3
San Jose, CA 95126
USA
andrei@angstromdesignautomation.com

ISBN 978-1-4419-6564-6 e-ISBN 978-1-4419-6565-3
DOI 10.1007/978-1-4419-6565-3
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2010929991

To our families

Preface

This Book and Simulation Software Bundle Project

Dear Reader, this book project brings to you a unique study tool for ESD protection solutions used in analog-integrated circuit (IC) design. Quick-start learning is combined with in-depth understanding for the whole spectrum of cross-disciplinary knowledge required to excel in the ESD field. The chapters cover technical material from elementary semiconductor structure and device levels up to complex analog circuit design examples and case studies.

The book project provides two different options for learning the material. The printed material can be studied as any regular technical textbook. At the same time, another option adds parallel exercise using the trial version of a complementary commercial simulation tool with prepared simulation examples.

Combination of the textbook material with numerical simulation experience presents a unique opportunity to gain a level of expertise that is hard to achieve otherwise. The book is bundled with simplified trial version of commercial mixed-mode simulation software from Angstrom Design Automation. The DECIMM^{TM} (Device Circuit Mixed-Mode) simulator tool and complementary to the book simulation examples can be downloaded from www.analogesd.com. The simulation examples prepared by the authors support the specific examples discussed across the book chapters.

A key idea behind this project is to provide an opportunity to not only study the book material but also gain a much deeper understanding of the subject by direct experience through practical simulation examples.

Each section of the book is accompanied by a set of simulation examples directly related to the main topics addressed in the section.

The examples are not just snapshots of the simulation results. Instead the reader has an option to use real simulation software tool. This allows the reader a practical opportunity to understand the ESD simulation examples by interactively studying the simulation results and changing simulation parameters within the limits of the trial version.

At the same time, the simulator software does not require any advanced skills in the technology computer sided design (TCAD) area. The authors of this book

and DECIMM^{TM} simulation tool developers believe that interactive features of the new tool will allow any electrical engineer or circuit designer to run the simulations successfully.

Although the free version of the simulation has some limitations compared to the full version, functionality of the free version is more than sufficient for it to be an indispensable tool in mastering of the subject of this book.

Nevertheless, the readers who do not want to take an advantage of the simulation or prefer to postpone the experience can read the textbook as any regular technical textbook.

The body of the book is prepared absolutely independent from the simulation examples, which are referred to only at the end of each chapter.

Subject and Purpose of This Book

ESD design for analog circuits is a very diverse and cross-disciplinary field. It involves an understanding of semiconductor device physics in strong non-linear operation regime deep knowledge of modern CMOS, BICMOS, and BCD process technologies, expertise in analog circuit design mixed with understanding of the product application conditions and specs, and even trends in marketing for analog integrated components.

Therefore, one of the major challenges accepted by the authors of this book consists in selection of an appropriate depth of material that will provide practical help in successful ESD design and can still be accessible enough to be used by a broad audience. This challenge requires both an appropriate simplification to fit the material within the limits of a single textbook and a new methodology to present the material. The methodology is based on combining the phenomenological approach and simulation results.

The book is organized in a hierarchy from the semiconductor device level to the product circuit level. In theory, each chapter can be studied independently. The seven chapters of the book address the following hierarchical levels:

I. Semiconductor Structures
II. Integrated Standard and ESD Devices
III. ESD Clamp Design Principles
IV. ESD Protection Network Design Principles
V. Protection of Signal Path Analog Integrated Circuits
VI. Protection of Power Management Analog Integrated Circuits
VII. System Level and Discrete Component ESD

This book targets all major aspects of ESD protection: device, network, and circuit design levels, mainly focusing on modern integrated components. System level and discrete component's ESD protection is addressed too in the last chapter.

This hierarchy is established in order to enable both sequential and independent study of the material depending of specific reader expertise and preference.

The authors expect that ESD engineers as well as students would appreciate a systematic representation of the material in increasing level of complexity. This approach covers the major background knowledge required for understanding the material.

Professionals already working in this field with engineering background expertise in device design may find it useful to skip the device chapters and proceed to the ESD network and analog circuit design material, while experts in circuit design may benefit from the device physics material that covers physical principles of conductivity modulation in semiconductor structures.

The authors believe that an option to interactively study the simulation examples will greatly benefit all readers.

The overall purpose of this book is to help professionals in the field to deal with the *Analog ESD Design* issues in their everyday professional work, attacking problems at all hierarchical levels starting from the device ESD level up to an implementation of integrated self-protecting solutions. The book is supposed to "arm" readers not only with important practical and technical knowledge, but also to add a complementary simulation experience that can be further developed with the light version of new industrial mixed mode simulation software DECIMM^{TM} from Angstrom Design Automation.

The Book Structure

The book is organized in increasing complexity of the discussed ESD subjects. This is not complexity in terms of understanding, since that will depend on the level of expertise – circuit or device design – with which an engineering professional is approaching this book. Thus, the level of complexity of the material across the chapters for circuit designers will perhaps be opposite to the level of complexity for device engineers.

The *Chapter 1* to the book below represents a short review of background material relevant to the ESD field, describing the authors' understanding of the field itself, ESD pulse specification and several other general aspects. At the same time, the introduction mainly refers the readers to other previously published books in the field, in order to maximize the space in the current book for material pertaining to the goal above. As well, the introduction establishes important definitions necessary for understanding the presented material.

Chapter 2 presents introductory material to the device design field. It provides a fundamental knowledge of conductivity modulation processes in the elementary semiconductor structures: p-n, p-i-n, n-p-n, p-n-p, p-n-p-n. The physical processes in these structures are described on a simplified phenomenological level that is easy to understand and further supported by rather simple independent simulation examples.

The material presented in *Chapter 2* about conductivity modulation in the elementary structures provides a necessary fundamental background both for understanding pulsed safe operating area (SOA) of the standard devices and for the snapback mode operation of the ESD devices. Chapter 3 covers both these aspects. It presents pulsed SOA for most typical integrated components in CMOS, BiCMOS, BCD, SOI, and SiGe process technologies, highlighting the parasitic devices formed in their structures. Pulsed SOA understanding is critical for the so-called ESD protection window realized for given circuit pins.

In parallel to each standard device, *Chapter 3* deals with most typical basic ESD protection devices that could be developed as "free" components in the given process technology. The important point of low-cost ESD design methodology is the creation of such "free" ESD devices. In case of "free" devices, it is assumed that the device can be obtained using only the available mask layers without violation of the minimum design rules exceeding physical capability of the process tools. Moreover, the "free" approach challenges self-aligned solutions and solutions that are closely based upon supported standard devices. In this case, reliability of the ESD devices can be linked to the corresponding supported device characteristics.

Chapter 4 targets the ESD clamp level using the ESD devices described in *Chapter 3* as building blocks for the variety of ESD protection clamps with desired characteristics according to pin functionality.

Chapter 5 presents a summary of ESD network design. It explores ways to apply different ESD clamp solutions to "assemble" an embedded ESD power circuit across all integrated circuit pins. The ESD network provides an ESD current path between different pin combinations that, in the case of analog circuits, may involve not only the ESD pad ring components but the internal functional circuit components as well. This chapter is a major reference point for practical ESD design of analog circuits that is further expanded on in *Chapters 6* and *7*, covering signal path and power analog circuits.

Chapter 6 is focused on ESD protection for signal path analog applications. It contains condensed introductory material that highlights design aspects specific to the ESD on a sub-block level. Signal path circuits and products are mainly represented in this chapter by high speed, precision and audio amplifiers, interface application, and digital–analog converters.

The scope of *Chapter 7* lies within different power management products. Analog ESD circuit design is discussed with the examples of dc–dc converters with integrated power devices, controllers, light management units, and LED drivers.

The final *Chapter 8* is focused on the system level of ESD design and ESD for discrete components. The goal of this chapter is not to describe the system itself, but instead the integrated components with pins interfacing directly with the system terminals. The major aspects are realization of the system level ESD protection on chip. Due to this being a relatively new topic for ESD design, this chapter contains a much more comprehensive introductory portion in comparison with the well described across literature sources introduction to ESD in *Chapter 2*.

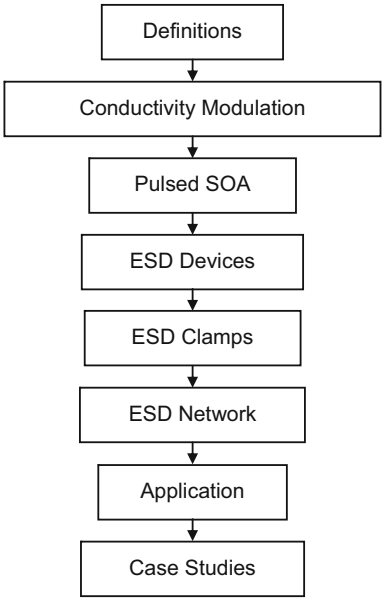


Fig. 1 Roadmap for the book composition

Finally, the last section of each Chapter provide a brief description of the simulation examples directly relevant to the material described in the Chapter. The examples are available for download from http://www.analogesd.com

Acknowledgments

The authors would like to acknowledge all their colleagues in the field who directly or indirectly helped with this book.

Vladislav Vashchenko would like to gratefully recognize his closest colleague and friend Ann Concannon from National Semiconductor for the many years of close collaboration in the ESD field that directly impacted and contributed to the creation of this book.

He would also like to acknowledge the contributions of his colleagues Alan Segerval, Donald Archer, Peter Hopper, Philippe Lindorfer, Alexei Sadovnikov, Lihui Wang, and David Lafonteese, as well as the many lead circuit designers and product engineers across the National Semiconductor Corporation with whom he worked on ESD projects with during the last decade. He is also deeply thankful for the multiple discussions of ESD subjects during the past 10 years with his colleagues in the field: Prof. Gaudenzio Meneghesso from the University of Padova, Prof. Elyse Rosenbaum from the University of Illinois, Robert Gauthier from IBM, Gianluca Boselli from Texas Instruments, Vesselin Vassilev from Novorel, Prof. Juin Lion from the University of Central Florida, Markus Mergens from QPX GmbH, Prof. Marise Bafleur from LAAS, Theo Smedes from NXP, Kai Goebel from AMES NASA, and many other colleagues and professional friends from the ESD Association and Industry. In particular, Vladislav considerably appreciates the many contributions and support of his current and past direct researcher collaborators and co-authors Dimitri Linten Mirko Scholz, Philippe Jansen, and Steven Thijs from IMEC, as well as PhD students Nicholas Olson, James Di Sarro, and Blerina Alija. He is also appreciative of his former mentors Prof. Vladimir Sinkevitch from Pulsar, Slava Osipov from AMES NASA and Boris Kerner from Daimler AG.

Andrei Shibkov would like to acknowledge many of his former colleagues from PDF solutions, in particular Carlo Guardiani.

The authors are grateful for analog design expert Vladislav Potanin, National Semiconductor, who reviewed Chapters 6 and 7 and provided very valuable comments to the material.

The authors greatly value the support and understanding provided by their families during the time devoted to this book project.

Finally, both authors would like to emphasize a critical contribution to this book by Yana Vashchenko, a first-year UCLA student, who spent many days during the

summer of 2009 performing a complete language and style edition for the whole book. They hope this practice will help her to become a better bioengineer in the future.

Contents

Chapter 1
Introduction

Readers proficient in the ESD field may decide to skip this introductory section. The section briefly summarizes the background directly relevant to the ESD field. This is done to facilitate the study of the following chapters for those who are not directly involved in the field or want to refresh the most important aspects of the knowledge. A broader spectrum of ESD background material is brilliantly covered in many books and reviews [1–7] written in the field, as well as in EOS/ESD Symposium Proceedings [8]. The purpose of this section is to summarize the general ESD approach to integrated components design and provide condensed reference material related to the ESD pulse specification, and standards. Finally, the most important definitions used across this book are established.

1.1 Analog and Digital in Prism of ESD Design

Historically, integrated circuits can be roughly subdivided into analog and digital categories. These classifications are currently used in the industry and are based upon functionality and design principles of ICs. Even in the case of complex mixed-signal circuits, different analog and digital domains can be identified and corresponding ESD methodology specific to either analog or digital circuits can be applied for each specific domain. The book title emphasizes "analog" as a major focus of the presented material due to several reasons. However, practically the material can be treated as universal ESD design guidelines for mixed-signal circuits, perhaps excluding practical examples of the most scaled down digital CMOS processes of 90–32 nm. At the same time, the digital domains in analog IC components built using 0.13–1 μm process technologies are well covered in Chapters 4 and 5 by active clamp solutions. These circuit blocks represent digital interface, digital control pins, and digital domains.

From an ESD design point of view, the digital circuits and digital circuit domains in mixed-mode circuits are different from analog circuits. The most straightforward way to identify the type of circuit is by the type of signal transferred through the circuit pins.

V.A. Vashchenko, A. Shibkov, *ESD Design for Analog Circuits*,
DOI 10.1007/978-1-4419-6565-3_1, © Springer Science+Business Media, LLC 2010

From the practical ESD design point of view, the difference between analog and digital designs is often reflected in the ESD pad ring circuit block. In case of the fully digital circuit, ESD protection is usually designed in the pad I/O (input/output) and power domains. In this case, the periphery of the digital circuit is expected to be practically fully isolated from the internal circuit.

When I/O and ESD library is designed and validated it can support wide variety of internal digital blocks with different functionalities. In this case, beyond the ESD/IO library creation itself, the major focus of ESD chip design is to ensure proper pad ring layout design that accounts for the voltage drop on metallization busses, number of clamps, and RC timers and make sure that current path for every pin-to-pin combination is addressed. As well, at appropriate voltage limitation, the current path is always expected to be confined within the ESD pad ring network. In the case of high-pin count digital ICs the problem complexity requires automated tools in form of ESD rule checkers.

The above is not intended to trivialize digital ESD design. Digital ESD design has its own complexity. For example, one of the most significant challenges is covering the specifics of high-pin count digital IC products, especially in the case of CMOS processes scaled down to 90–32 nm gate dimension. In this case, one of the major challenges is CDM (charged device model) pulse protection of the large form factor packages. There are many challenges in protection of high-speed and RF I/O pins, as well as system-level protection.

However, in the case of widely used analog 0.5 μm process technologies the digital pin protection hardly presents any real challenge. The methodologies for such protection are very well established.

The major reason for discussion of digital vs. analog circuitry is to emphasize the specific of ESD protection approach and solutions.

In the case of digital design, one can expect that the ESD network building blocks in pad ring design will typically provide only a single connection to the internal circuit. This connection can be relatively easy to analyze in the ESD current path analysis. At the same time, latch-up isolation is automatically provided by ESD library solutions with no or minimized interaction of the ESD and I/O circuit with other circuit nodes due to appropriately designed guard rings.

In opposite, analog ESD design guarantees no such condition. In general, the analog ESD protection design can be expected to account for the fact that the analog pad might have multiple connections to the internal circuit node. The connections can be realized both directly and indirectly by the coupling through the power components, for example, large drain–gate capacitance of multimillimeter width NLDMOS power array.

Moreover, most of these internal circuit nodes will generally have an unknown transient bias and current conditions for different pin-to-pin zap combinations. This makes it rather difficult to guess at what point of safe operating area the active device connected to the pin is. During ESD pulse, power components of the analog circuit may switch to on-state during ESD event and conduct a substantial amount of current during a part of the ESD pulse. A “sneak” current path can be formed

between the ESD clamp and internal circuit components depending on the layout of the circuit, especially in case of high-voltage device.

We would like to use this type of differentiation to underscore the essence of separation between the digital and the analog *ESD designs*. The fact that a substantial amount of analog products can contain digital interface pins and domains does not change this approach to ESD design principles.

Thus, in the case of *digital ESD design*, the ESD approach can be unified and formalized to be somewhat independent from the internal circuit blocks as long as their type is identified and appropriate ESD or ESD/IO library cells selected with ESD pad ring and created in accordance with the library guidelines. In opposite, in the case of *analog ESD design*, a practical ESD design should often be customized for new circuit changes, taking into account possible alternative scenarios for ESD current conduction through the internal circuit. These potentially new scenarios should be then taken into account for both ESD network design and ESD clamp choice.

Thus, one of the most critical features of the analog vs. digital designs is the ESD protection network and the internal circuit in general cannot be separated. This creates a need for the ESD engineer to understand the analog circuit at much greater depth.

Another important distinct feature from ESD perspective is the product pin count. Digital products often have hundreds of pins, while some small form factor analog products might have as few as three to five pins. This fact automatically brings into consideration the critical issue of the space used on the chip for ESD protection.

In the case of digital circuit with a high pin count, the chip periphery is relatively large and a distributed active clamp solution is relatively spatially optimal, usually requiring space, the size of one pad at each pad. Therefore, perhaps over 90% of the digital circuits are protected by the distributed active clamp solution.

A different situation can be often found in case of analog circuits. There are many examples of small-pin-count analog ICs where over 50% of the silicon die space is taken up by ESD clamps. In this case, space saving and small footprint solutions can provide a major impact on the product cost.

At the same time, the voltage tolerance for the ESD clamps that corresponds to the digital signal levels is rather low. In the case of mature process technologies, this makes even the local protection option relatively easy to address. The opposite situation is found in case of high-voltage analog products where an optimal high-voltage solution might be rather hard to find without changes to process technology.

The purpose of this book is to help professionals in the field to deal with the *analog ESD design* issues in their everyday professional work, attacking problems at all hierarchical levels starting from the device ESD level up to an implementation of integrated self-protecting solutions. The book is supposed to "arm" readers not only with important practical and technical knowledge but also to add a complementary simulation experience that can be further developed with the new mixed-mode simulation software DECIMM^{TM} from Angstrom Design Automation.

1.2 Important Definitions

1.2.1 ESD Protection Network

Practically every book written in the ESD field well describes the nature of phenomena behind ESD charge accumulation and the discharge events [1–4]. Over time the specs created for industrial application move further away from real events that may cause charging or the objects that can provide discharge through IC pins.

Therefore, to avoid redundancy and save space in this book, the starting point for the ESD subject is presented from a slightly different angle.

Without loss of generality, we will assume the following approach to the subject. An integrated circuit represents an object that contains both internal circuitry components and external pads. A discrete component can be treated as a particular case of IC where internal circuitry is represented by a single or few devices packaged together. Similarly a system can be considered as a combination of circuit blocks that contain mounted, integrated, and discrete components with external terminals.

The pads are bonded to package and pins. The pins could be package pins, system socket connectors, or device leads.

The ESD protection task will be further treated simply as an added ability to the integrated components or systems, or discrete components to withstand certain ESD pulse specification.

In this case, the ESD protection capability is treated as a part ICs or system specification similar to other normal specification parameters for the internal circuit performance and reliability.

In most practical cases of analog integrated components design, this added ability to withstand certain level of ESD pulse spec is achieved through a co-design of the internal circuitry and the ESD protection network. Therefore, an idea of separation between the ESD network and internal circuitry as well as corresponding requirements for non-conflicting existence of the ESD protection solution with the circuit performance is perhaps irrelevant.

Nevertheless, the ESD co-design goal is to minimize the impact of the ESD network components on the ideal circuit performance that could have been accomplished if ESD performance was not specified.

Similar reasoning is relevant for system-level ESD where normal product operation during ESD and electrical overstress (EOS) events may be required. System-level ESD protection is presented in Chapter 8.

Thus, a problem of ESD development can be formulated as the implementation of an embedded capability of integrated products toward withstanding standard ESD tests according to defined specification.

From a circuit design point of view, this essentially means an extra functionality of the integrated product circuit at some specific high-current pulse conditions in addition to normal operation specification targets.

There are several ways to realize such capability. One of the major principles is to implement additional peripheral ESD protection networks connected in parallel with the original functional circuit blocks. However, analog ESD design often relies

on the self-protection capability of the internal components. The self-protection of the internal components can be exploited both up to the level of the full ESD current conduction and using much smaller current levels through the internal circuit to enable a two-stage protection. In principle, a pin-specific combination of the above measures is usually used.

Thus, in general, the product can be considered as a superposition of internal functional circuit blocks and additional pulsed power circuit dealing with ESD current. This circuit can partly use the components of the functional circuit blocks. This pulsed power circuit can further be understood as the *ESD protection network*.

The major functionality of the ESD protection network is to provide a high-current path with appropriate voltage limitation in case of ESD discharge applied for every pin-to-pin combination. Under the appropriate voltage limitation requirements, a limitation of the voltage in the ESD time domain below the pulsed safe-operating area (SOA) is understood for the device. Thus, the ESD protection network should not only provide the ESD current path but also limit the voltage below absolute maximum rating conditions realized at each pin. The pulsed absolute maximum rating provided by the circuit at each pin is rather complex figure of merit that depends on time domain, rise time, maximum ratings of the devices connected to the pin, coupling of their control electrodes, and other factors that could be related to a parasitic current path realized in the actual layout.

Thus, the task can be converged into co-design and embedding of the ESD network. ESD network design principles are discussed in Chapter 5.

As has been already mentioned, this network can partly be realized using the self-protection capability of the active circuit components. However, major building blocks are the *ESD protection clamps,* connected by an appropriate metallization routing for given pulsed current level.

An ESD protection clamp usually presents itself as self-triggered in the high-current state. The self-turn-off event of clamps usually occurs due to discharge of the ESD pulse. There are two major categories of ESD protection clamps. The first uses the RC network to control on and off state conditions for the active devices. In this case, the high-current device is fully controlled in high-current mode and can be turned off by the RC network.

Another major class of *ESD clamps* is based upon avalanche breakdown and snapback *ESD devices* that involve different conductivity modulation mechanisms in order to achieve high-current conditions. Unlike active clamps, the turn-off of ESD devices with conductivity modulation usually cannot be practically controlled by a driver circuitry. The ESD device operates in the conductivity modulation mode up until the voltage conditions are changed simply due to the end of the ESD discharge, or the voltage level will be naturally reduced below the holding voltage of the snapback components or below the breakdown voltage.

The major principles for ESD clamps design are discussed in Chapter 4. The ESD devices are presented in Chapter 3. The principles of conductivity modulation are presented in Chapter 2.

1.2.2 ESD Clamps

The building blocks of the ESD protection network are the *ESD clamps*. Under ESD clamp we will further understand a simple circuit that provides pulsed ESD current path under certain conditions. These conditions are generally defined as achieving the critical voltage level or a fast rise time. In most sophisticated cases, the clamp operation can be controlled or enabled by the additional control electrode. ESD clamps are discussed in Chapter 4.

The clamp can be designed using integrated components operating in normal operation mode of monopolar or bipolar current conduction. This principle is the basis of active clamps discussed in Chapter 4.

One of the most critical requirements of the ESD protection network toward competitive advantage of the product is to occupy the feasibly smallest space on the chip. Therefore, a proper way to achieve such capability is the implementation of small footprint device-level solutions, taking an advantage of several isothermal conductivity modulation mechanisms that can be realized in such devices. Thus, for analog design, the most useful clamps are the ones reversibly operating in high injection and breakdown modes.

Typical device-level ESD solutions are essentially pulsed power devices with some biasing components that are specifically designed to work in the high injection, breakdown, and conductivity modulation modes.

ESD clamp development involves a rather broad set of cross-disciplinary tasks in the fields of circuit design, physics of semiconductor devices, physical ESD clamp design with non-linear physics aspects of the operation, effects in interconnects, materials, and topological array problems to balance the current density distribution. Finally, the most important necessary expertise involves a deep understanding of process technology including the options and the limits.

Meantime, as has been shown in [9] and will be particularly illustrated in detail in Chapter 2, the physical phenomena responsible for conductivity modulation used in ESD devices are very limited to several physical mechanisms that can be analyzed using elementary diode, triode, and thyristor semiconductor structures. These practically useful mechanisms include avalanche breakdown, avalanche–injection, double-avalanche–injection, and double injection. At this point of ESD field evolution, there were no reports of using other conductivity modulation mechanisms for practical ESD design. Similarly, due to short ESD pulse time domain, the conductivity modulation mechanism based upon thermal carrier generation plays a secondary role, mainly limiting the high current capabilities of the ESD device itself [9]. This limitation is discussed in Chapter 3.

The operation in these breakdown conditions generates high current densities presented in Table 1.1. In case of avalanche breakdown, the carriers are generated in the region with a rather high electric field and are separated proving practically no mutual space charge neutralization. In this case, relatively low current density can be achieved. In case of avalanche–injection, the mutual space charge compensation of the carriers generated in the avalanche region and injected creates more favorable conditions for high current density. The highest current levels are provided in

Table 1.1 Conductivity modulation mechanisms realized in ESD devices for high current density

Conductivity modulation mechanism	Typical ESD devices	Typical lateral current density (mA/μm)
Avalanche breakdown	Avalanche diodes; blocking junctions, PMOS, PNP	0.01–0.1
Avalanche–injection	Snapback NMOS; NPN, field oxide devices	0.1–3
Double-avalanche–injection	P–i–n, M–i–n diodes	0.1
Double injection	LVTSCR, SCR, bipolar SCR, LDMOS-SCR	10–100

the case of double injection, when the avalanche multiplication is supressed in the structure (see Chapter 2).

One of the challenges in creating an appropriate ESD device is balancing the conductivity modulation current density inside the device to achieve an appropriate current level. In this case, a linear width scaling of the ESD current can be expected in the device. Since some of the mechanisms provide a positive feedback that in general can result in uncontrollable current density increase, a negative feedback loop should be implemented in the device to limit the current density below critical limits.

ESD devices are discussed in Chapter 3, while basic principles of the conductivity modulation in semiconductor structures are discussed in Chapter 2.

1.2.3 Absolute Maximum Limits and Pulsed SOA

As has been stated above, on a formal circuit design level upon given spec, the ESD functionality is achieved by implementation of a pulsed power circuit embedded into the normal circuit using shared and dedicated components. This partly virtual secondary pulsed power circuit should not only provide a pulsed power operation for the pin-to-pin current path scenario but also limit voltage below the damage level of the internal circuit blocks.

Thus, realization of the voltage waveform parameters in the ESD time domain is a major challenge, considering the absence of the limitations in normal circuit operation regime. Thus, the ESD protection network should provide voltage waveforms that will limit the voltage below the absolute maximum limits on the pin, but at the same time will not cause false turn-on during normal operation. Practically, this means that the turn-on voltage should be realized above absolute maximum limits specified for the circuit operation modes in all data sheet conditions including the temperature range.

These two limits define the so-called ESD protection window. While the lower limit of ESD protection window can be sourced from the circuit application spec, the upper limit is a more complex matter. In principle, a preliminary idea about

absolute maximum limits for the particular pin can be "extracted" from the pulsed SOA of the devices directly connected to the pin. However, the problem is that most of electrical design rules for process technology usually do not provide pulsed SOA in the ESD time domain.

Another major issue is unknown coupling effect of the control electrodes of the devices connected to the pin in the ESD test modes. In this case, the uncertainty is related to the selection of the SOA regime for identifying the ESD protection window.

There are no universal recipes for identifying the ESD protection window for analog circuit pins before the actual ESD tests. One of the most productive approaches is based upon transmission line measurements (TLP) for pulsed SOA. These SOA aspects are discussed in Chapter 3 for most typical devices realized in BCD (bipolar CMOS DMOS) process technology.

1.2.4 ESD Pulse Specification

During the era of integrated product development, various specifications for ESD pulse have been created and became industry or custom standards. The detailed description of the ESD pulse characteristics and the equivalent circuits can be found in the introductory chapters of numerous popular books in the field, for example [1–4], original publications [8], and standard documents, for example [10–13].

The standards for the human body model (HBM) pulse are as follows: ESDA (ANSI) STM5.1-2001; JEDEC JESD22-A114-E; IEC 613240-3-1; AEC Q100-002 REV-D; and EIAJ ED-4701/304. Similar standards can be found for the machine model (MM), the charged device model (CDM), and the ESD system level [13]. The consolidated summary of most commonly used ESD pulses is presented in Table 1.2. Standard non-system-level corporate requirements for the HBM, MM, and CDM ESD passing levels are 2 kV, 200 V, and 1 kV, respectively.

Table 1.2 The most common examples of non-system-level ESD pulse parameters

ESD pulse	Peak current to pulse voltage ratio (A/kV)	Rise time/pulse width (ns)
Human body model	0.67	2–10/150
Machine model	17.5	2–10/66–90
Charged device model	9	0.25/2
EN (IEC) 61000-4-2	3.75	0.2/50–150

From an ESD device design challenges perspective, there are two major spec types: the non-system level and the system level. The non-system-level spec usually targets some minimal requirements to protect the integrated circuit component while incorporating it into the system, packaging, handling, and electrical tests. For example, the HBM ESD pulse spec for the non-system level requires withstanding

some pulsed current level of ~1.33 A with the rise time of ~2–10 ns (Table 1.2). Respectively, the test is conducted in the conditions of the unpowered circuit.

In opposite to the non-system packaged specs, system-level specs are targeting protection of some circuit pins under normal operation conditions. In addition to this, usual system-level requirements target much higher current levels that can be practically realized in a non-ESD protected environment. The complexity of the system-level protection problem is related to a possibility of transient latch-up. Transient latch-up can be realized in case if ESD clamp provides a holding voltage lower than the power supply voltage under the minimum holding current below the current that can be provided by the power supply.

Due to fast rise time, in most cases ESD pulse automatically provides the conditions for pure electrical turn-on. Electrical current instability is initiated in quasi-isothermal conditions and provides further triggering of a high-current conductivity modulation state. In most practical cases, the lattice temperature change can be neglected before the triggering due to uniform current distribution and short time before the triggering.

After the switching, heat dissipation becomes significant. However, the heat dissipation scenario is significantly different from dc operation. Due to rather short pulse duration, the heat dissipation is realized in a rather small area of few microns in the vicinity of the device's active region. Thermal heat dissipation (as well as electro-mechanical stress and dielectric breakdown) and backend limits provide physical limitations for ESD device operation. These effects are illustrated in Chapter 3.

1.2.5 Breakdown and Instability

In the majority of cases, ESD events in semiconductor structure are close to adiabatic conditions. In spite of high current density, due to very fast switching time and fast ESD pulse (~100 ns), the heat generation is confined in rather local area (Fig. 1.1).

This phenomenon eliminates a significant amount of physical effects related to the carrier generation in the drift regions, epi-layers, substrate, and corresponding current instability phenomena.

The thermal effects further play an important role in the final irreversible catastrophic phenomena when ESD device fails.

The phenomena of irreversible breakdown, burnout, thermal and isothermal instability, and current filamentation are often mentioned in respect to semiconductor device failure in the case of both ESD and EOS events.

In this book, the notion "breakdown" preserves its elementary primary sense. The breakdown is treated similar to the effect in p–n semiconductor junctions as a process of sharp current increase that is caused by the carrier generation current.

Practically for the ESD field, the only important cases are isothermal avalanche and tunneling breakdown.

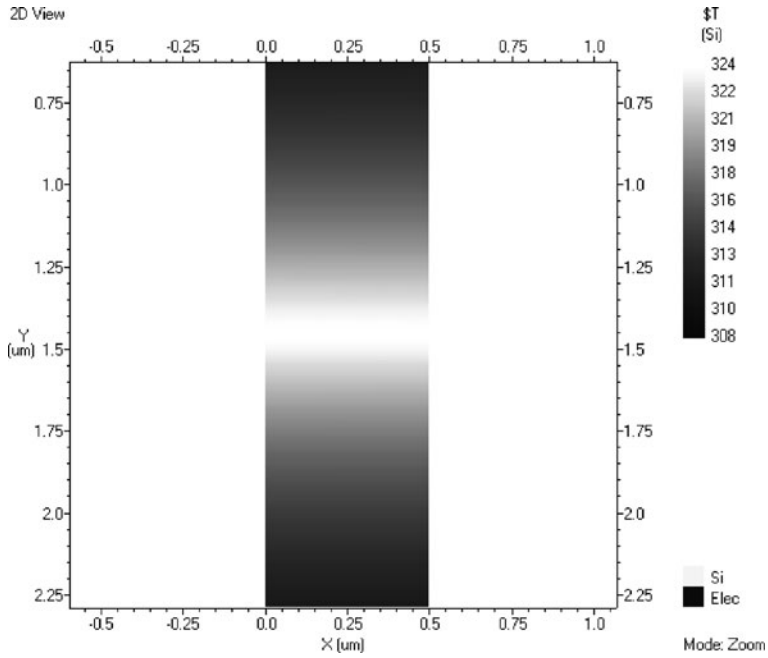


Fig. 1.1 Simulation example: local heating in collector region of NPN transistor caused by 100 ns stress pulse

In avalanche breakdown of the reverse-biased p–n junction, current I increases with voltage U increase according to the known empirical formula [14]:

$$I \sim \frac{1}{1 - \left(U/U_{\mathrm{BR}}\right)^{n}},$$

where $n = 4$–6 for Si material and U_{BR} is the avalanche breakdown voltage. The dependence $I(U)$ in this case of "classical" avalanche breakdown is strictly monotonic.

A different kind of breakdown could be observed in a case where injection and conductivity modulation are involved. In this case, current increase can create complex S-shaped I–V characteristics of the devices.

The current instability can be defined as a process of uncontrollable current increase due to voltage decrease.

This phenomenon in general may or may not result in irreversible device failure due to the appearance of some limiting factors. In the case of the ESD device, the current instability is used to create low dissipated power conditions in conductivity modulation mode. To avoid high-amplitude filaments, the local current density is limited internally on the device. Implementation of such limitation is an essential part of ESD device architectural design.

Thus, the physical definition of the breakdown is just a sharp current increase under positive differential conductivity. In other words, the breakdown itself does not include any positive feedback. The major breakdown mechanisms in semiconductors are the avalanche and the thermal, although several additional mechanisms can be found, for example, the breakdown related to the change in trap charge state or dielectric breakdown.

In opposite to the breakdown, different electrical and thermal instabilities [9] include a positive feedback. The thermoelectrical instability phenomena in semiconductor devices are rather complex. To provide a "quick start" in understanding the physical sense of these phenomena, we present below the example of thermal instability in semiconductor structure.

The best example is the avalanche–injection conductivity modulation in NPN structures discussed in Chapter 2. An accurate analytical description for conductivity modulation mechanisms is rather complex. An example of the analytical description for the case of thermal breakdown is used below to explain the current instability phenomenon itself. Often this thermal instability case can be responsible for the physical limitation of the current level provided by ESD device [9].

In the example of bulk semiconductor structure, if at a constant voltage $U << U_{BR}$ the devices are heated by some external current source, then from some temperature level the current through the sample will grow sharply according to an exponential dependence $I \sim \exp\left(-E_G/kT\right)$, where E_G, T, and k are the energy of band gap, lattice temperature of semiconductor material, and Boltzmann's constant, respectively. This thermogeneration process for carriers is considered as a thermal breakdown.

It is assumed that on a uniform sample of bulk semiconductor with length l and area S the voltage U is supplied. The current density j through the sample is equal to $j = \sigma E = \sigma U/l$, where E is the electric field in the sample and σ is the conductivity of semiconductor material. The conductivity of semiconductor material in this case can be expressed by $\sigma \sim \exp\left(-E_G/kT\right)$.

Then, the generated heat per volume unit is equal to σE^2. It is also further assumed that heat dissipation is provided by exchange with ambient space of fixed temperature T_S. Then, heat dissipation from the surface will be proportional to $(T - T_S)^\beta$, where T is the temperature of semiconductor region and $\beta = 1$–2. The heat balance equation is given by

$$\exp\left(-\frac{E_G}{kT}\right)\frac{U^2}{l^2} = K\,(T - T_S)^\beta\,, \tag{1.1}$$

where the proportionality factor K depends on physical and geometrical parameters of the device structure. In a certain range of values of U this equation can have two solutions: T_1 and T_2 (T_1, $T_2 > T_S$). The solutions correspond to two different current values I_1 and I_2.

The I–V characteristic of such a structure is no longer as simple as avalanche breakdown and has an S-shaped view (Fig. 1.2b).

The physical meaning of the I–V characteristic can be explained as follows. On the initial region "OC," the current increase obeys dependence close to Ohm's law.

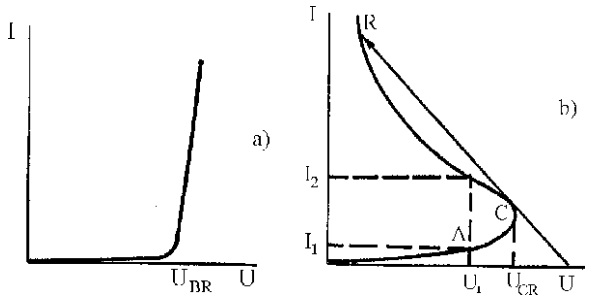


Fig. 1.2 *I–V* characteristic at avalanche breakdown of p–n junction (**a**) and at the thermal instabilities in semiconductor resistor (**b**)

An appreciable deviation from linear dependence begins when the heating reaches a higher level. The conductivity increase is connected with intensive thermogeneration of electrons and holes in the sample. Insignificant increase of the voltage on the thermal breakdown region "CA" results in a sharp increase of heat generation.

Up to a certain limit, the heat generation is balanced by an increase in temperature since the heat dissipation is proportional to $(T - T_S)^{\beta}$. However, at some $U > U_{CR}$, the exponential increase of the heat generation can no longer be compensated by heat dissipation. In this state, (1.1) has no solutions and the semiconductor sample has no stationary states, respectively. This means that in the voltage source regime at $U > U_{CR}$, a sample will be uncontrollably self-heated up to its destruction.

A similar loss of thermal stability or an uncontrollable process of transition into a new state usually means instability. In this case, the instability is of a thermal nature. From Fig. 1.2b, the stable states of a sample in the case of thermal instability can be achieved only at corresponding voltage decrease *U*. If the circuit provides a sufficient load resistance, then the thermal instability may finally evolve into a stable state that corresponds to an *I–V* characteristic with negative differential conductivity (NDC) (Fig. 1.2b, state R).

The load characteristic CR in the case of such a device in circuit operation could be called snapback.

Current instability in real device structures is not always easy to interpret due to an additional spatial current instability phenomenon.

Since in the case of current instability the local current density dependence upon voltage has negative differential resistance, a typical consequence is current stratification into filaments and hot spots.

These phenomena may or may not damage the device depending on the local damping implemented in the ESD device.

This damping can often act up to a certain electrical power limit followed by the hot spot formation in the silicon surface. Usually, the hot spot is a narrow region with a dimension of a few micrometers with concentrated current and elevated temperature. In bipolar transistors, similar current localization results in thermal breakdown. In this case, a sharp temperature increase may result in metallization melting.

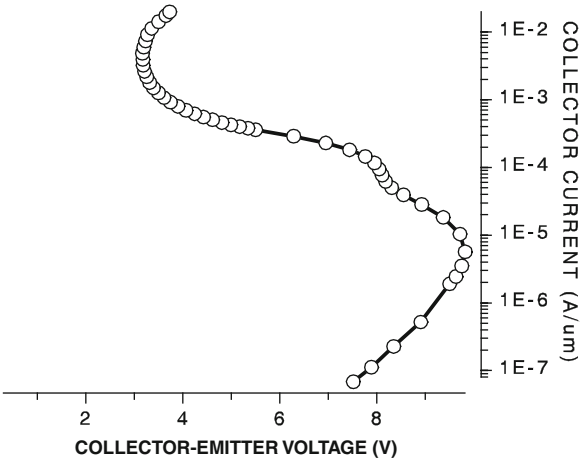


Fig. 1.3 Simulation example: snapback *I–V* in NPN

The example of thermal instability is provided above to deliver a primary understanding of processes that are extremely widespread in real device structures.

Alternatively to the analytical methods the thermal instability for given NPN structure can be obtained by numerical simulation results (Fig. 1.3) thus enabling additional way of learning.

A few summary points emphasized at the end of the introduction.

In opposite to digital, the analog circuits or analog domains in mixed-mode circuits are rather diverse in terms of the voltage tolerance and signal spec and often require small pin count. In this condition, practically every pad may require a separate ESD protection clamp limited by rather critical small footprint requirements.

This task could be solved in most cases only on the device level and thus requires understanding in three major key areas:

(i) The principles of operation of ESD devices operating in the breakdown and conductivity modulation conditions (secondary breakdown)
(ii) ESD clamp and ESD network design based upon these devices
(iii) Application of the ESD network to analog circuits and network–ESD circuit interaction

These key areas are addressed step by step in this book to enable a proficient, practical ESD design for those who are involved in the field.

Of course, once this understanding is established, a more detailed experience in the field is required based upon real product case studies. Nevertheless, we believe that in-depth knowledge about ESD components, networks, and practical examples for the analog circuit ESD protection (Chapters 6, 7, and 8) should bring ESD expertise to a new useful level and will be a relevant contribution to the field of ESD.

DECIMM^{TM} Simulation Examples for Introduction

To download a trial version of the numerical simulation software and request an electronic license key please visit http://www.analogesd.com
To download libraries with simulation examples for this chapter please visit http://www.analogesd.com/Chapter1.html
List of examples is subject to change.

Example 1.1 Simulation of the avalanche breakdown I–V in elementary quasi-1D structures

Example 1.2 Simulation of the S-shaped current instability in 2D NPN device

Example 1.3 Simulation of the transient thermal-coupled mixed-mode solution for 2D NPN, demonstrating hot spot area during ESD pulse confined within 1 μm

Chapter 2
Conductivity Modulation in Semiconductor Structures Under Breakdown and Injection

2.1 Important Definitions and Limitations

Understanding semiconductor structure operation under ESD pulse conditions at the physical level is critical for successful protection circuit design. In spite of the use of a variety of ESD protection devices and clamps for analog circuit protection, there are several fundamental physical effects taking place during a high-current ESD event. These effects are discussed in this chapter.

At high-current operation in short-pulse conditions and under ESD pulse in particular, semiconductor devices obey certain principles. These principles are confined within a rather limited set of basic isothermal conductivity modulation mechanisms [9]. Physical limitation of both the electrical safe operating area (SOA) in standard (provided by the process technology) devices and the ESD snapback devices is the result of the same fundamental principles.

This introductory chapter provides a quick start summary required for understanding the most important conductivity modulation effects utilized on the device level.

2.1.1 Basic Semiconductor Structures

Implementation of ESD devices in the given semiconductor process technology usually involves significant complexity. This complexity is the result of combination of different factors, for example, multiple complex implant profiles, multiple regions and interfaces, topology, and 3D effects. In some cases, several alternative ESD current paths can be found in real ESD cells depending on the current level. These paths can change, depending on the transient conditions specific to given ESD pulse type and cell topology. One of the most typical examples of this occurrence is in the NPN BJT device in BCD process technology, where competition between the vertical and the lateral current transport can be observed at a typical combination of structure parameters.

Nevertheless, based upon present understanding [9], it can be concluded that all these processes can be identified within in a very limited number of avalanche–

V.A. Vashchenko, A. Shibkov, *ESD Design for Analog Circuits*,
DOI 10.1007/978-1-4419-6565-3_2, © Springer Science+Business Media, LLC 2010

injection effects. Moreover, most of these effects can be treated as adiabatic due to a relatively short ESD pulse duration with the generated heat dissipated within very small region ($\sim$1 μm). This condition significantly simplifies analysis of the ESD device operation.

In spite of a superposition or a "domino" effect observed in a real life scenario, the avalanche–injection effects can be understood based upon a type of conductivity modulation. The type of conductivity modulation can be linked to a conductivity modulation in one of the elementary semiconductor structures. These structures are analyzed in this chapter.

Thus, the major goal of this chapter is to provide an understanding of current conduction in elementary structures under avalanche–injection conditions. This understanding is directly useful for analysis of the processes in real ESD devices and clamps in pulsed operation mode.

Since this book is written in the field of analog IC design, the major focus is confined to Si and Si–Ge semiconductor materials used in today's microelectronics for analog circuits. However, similar principles can be applied to compound semiconductor materials GaAs and GaN [15–22].

Due to short-pulse conditions, certain assumptions are made about the electrical regime in ESD time domain. In particular, most of the phenomena can be treated as an isothermal case. Another important assumption is silicon semiconductor material parameters. Under these assumptions there are only five elementary semiconductor structures that can provide a practically relevant current density level under conditions of breakdown and injection. These elementary or primitive structures are as follows:

(i) *p–n structure* operating at reverse bias in avalanche breakdown mode
(ii) *n–p–n structure* operating in avalanche–injection conditions
(iii) *p–n–p structure* operating in avalanche–injection conditions
(iv) *p–n–p–n structure* operating in double-injection conditions
(v) *p–i–n structure* operating in double-avalanche–injection conditions

To stay within practicality, these five elementary structures are presented and analyzed with heavily doped contact regions of type p^+ or n^+. Respectively, the internal p, i, or n-base or drift regions are assumed to have significantly lower doping, in comparison with the contact regions. Consequently, the conductivity modulation analysis is accomplished within the area outside of the contact regions which remain quasi-neutral, i.e., the dominant part of the space charge region is confined within the bulk of the lower doped drift regions.

Thus, in reference to p–n structures, the structures p–n–n^+, p^+–p–n^+, or p^+–p–n–n^+ are assumed to be practically relevant cases, where the acceptor and donor levels in the corresponding regions are $N_{Ap+} \sim N_{Dn+} >> N_{Ap}, N_{Dn}$.

There are several other structures that are potentially useful for ESD. For example, Schottky diodes operating in avalanche–injection conditions, tunneling (Zener diode) structure, discharge gaps, polymer suppressors. However, these devices are

not discussed in this book due to a particular focus on the most practical analog ESD design applicable to the integrated circuits.

2.1.2 Conductivity Modulation and Negative Differential Resistance

Under *conductivity modulation* in the semiconductor device structure we will simply understand a non-linear behavior of the structure that results in a significant change of conductivity due to voltage or current applied to the structure contact regions. This is a definition that is based upon the external structure parameters. Similarly, conductivity modulation can be defined through internal structure parameters as a change in the conduction properties of particular structure regions as a result of change in carrier balance due to avalanche and injection processes. The conductivity modulation in multiterminal structures is the result of an internal avalanche–injection process initiated by change in the applied current or voltage. This is emphasized in opposite to the change of the structure conductivity due to injection/extraction of carriers from the base contact region or creation accumulation/depletion regions by the field control electrode.

From a practical ESD device design point of view, the most interesting cases are related to a non-linear change of the conductivity that results in negative differential resistance effect. The negative differential resistance can be simply defined as a decrease in the voltage drop on the structure contacts with a current increase through the contacts. This effect is usually observed as a formation of S-shaped *I–V* characteristics, although in some cases the saturation resistance of the drift and contact regions compensates the internal negative differential resistance.

A different known type of non-linearity is the negative differential conductance that corresponds to formation of N-type *I–V* characteristics. In this case, the voltage increase results in current decrease. This phenomenon is practically observed under formation of the traveling Gunn domain in GaAs [9] or in case of trap charge change. However, it appears that at this point this effect is barely useful for ESD protection solution design.

In opposite, ESD applications of the negative differential resistance effect are extremely useful. This physical effect provides a power-efficient way to dissipate ESD current under a lower voltage drop on the protected pin. As a result, the lower dissipated power on the ESD device both provides an opportunity for a small footprint solution design and simultaneously enables realization of the desired voltage waveform during ESD pulse. The last is achieved by the low clamping voltage that guarantees survival of the internal circuit components while keeping the electrical pulsed regime within the absolute maximum ratings.

In the case of a real circuit, full S-shaped *I–V* characteristics are hard to observe during the negative differential resistance (NDR) effects due to a finite load resistance. Therefore, a part of NDR might be represented by the load resistance. For example, in the case of a 50 Ω transmission line pulse (TLP) system [2], a part of

the S-shaped *I–V* characteristic will be created with a corresponding 50 Ω load resistance. A 1.5 kΩ HBM waveform tester may enable observation of more details on the S-shaped *I–V* characteristics, even demonstrating partly real negative differential resistance of the structure itself.

Thus, in the case of an operation with finite small load resistance, the device achieves negative differential resistance conditions at some critical voltage, causing a self-turn-on or self-triggering into the corresponding high-current state. This phenomenon is usually referred to as a snapback among ESD application engineers and circuit designers. This term will further be used across this book to identify the associated phenomena described above.

At the same time, according to Vashchenko and Sinkevitch [9], as it will be demonstrated below, the real physical processes responsible for such behavior represent a very limited set of mechanisms based upon a positive feedback loop between avalanche and injection phenomena.

Beyond the practical use of negative differential resistance in ESD devices to provide ESD current discharge and desired voltage waveforms, this phenomenon has a much more general nature. Practically, as shown in [9], together with thermal conductivity modulation effect, this phenomenon is responsible for physical limitation of electrical regimes of all real structures. In particular, it is essentially responsible for physical limitation of the ESD device operation in pulsed conditions. These aspects will be addressed in Chapter 3.

2.1.3 Spatial Current Instability, Filamentation, and Suppression

Negative differential resistance is often linked in literature with the current instability and spatial current instability phenomena. Indeed, on the circuit level, the current instability in the structure can be intuitively understood as an uncontrollable current increase when some critical voltage level is applied to the structure. Specifically, such an effect can be observed during snapback depending on the load resistance.

Most of the ESD pulses produce a current level of ∼1–10 A, while most of the structures are capable of providing current densities of ∼0.1–10 mA/μm. Therefore, the discussion of ESD devices that can be used in the clamps automatically involves largely distributed objects designed as multifinger devices and arrays.

At the same time, one of the important non-linear physical effects observed in distributed structures with structure-level positive feedback is directly related to a different type of current instability [9, 23–27]. This type of instability is the instability of the uniform current density distribution across the structure width. Thus, in the case of negative differential resistance, the current distribution in the sufficiently wide semiconductor structure is in general spatially unstable and results in the formation of current crowding, current redistribution, and current filamentation effects.

The spatial current instability evolution scenario depends upon the electrical regime and load characteristics. However, the typical spatial dimensions for

the structure's positive and negative feedback processes determine the possible solutions for non-uniform current states. Spatial instability of the current density distribution may result in the formation of rather complex non-uniform states with a current density that may significantly exceed the initial uniform current density [26]. A simplistic example of such a state is the isothermal current filament.

Formation of spatially non-uniform states in distributed semiconductor structures is one of the interesting phenomena of non-linear physics [26, 28, 29]. In this chapter, numerical simulation results will be presented for several elementary structures in Section 2.8 in order to demonstrate the general physical principle behind the phenomenon.

One of the critical targets for successful design of the ESD device with certain positive feedback is to implement structure-level negative feedback. This negative feedback needs to be engaged at a certain current level and limit the current density to below the safe conditions. For example, in grounded gate snapback NMOS, discussed below, the positive feedback realized due to avalanche–injection conductivity modulation in the parasitic n–p–n structure is compensated on the structure level by the negative feedback provided by current saturation in the drain ballasting region. This device will be discussed in detail in Chapter 3.

In general, the negative differential resistance effect leads to the spatial current instability in the device. The spatial current instability results in current filamentation. The filamentation itself is not necessarily an effect that immediately results in irreversible damage of the device. Moreover, the snapback NMOS at high ESD current represents an operation mode in which the device supports the current filament regime visible at a small current level. In this case the filament amplitude is limited by the drain ballasting region below the safe limits in pulsed regime. At the maximum level of current, the operation mode practically corresponds to a filament mode with a width equal to the structure width. At the same time, burnout of such a device is the result of conductivity modulation of the ballasting region that in its own turn results in much greater amplitude filament that usually almost immediately locally melts the device.

Thus, while in an appropriate ESD device the peak current density is limited, the pulsed operation of the ESD device is expected to be fully reversible. An opposite situation is realized in a standard device. With some exceptions, the standard devices for the given integrated process will not provide a reversible operation in negative differential resistance conditions and following snapback mode of operation. Therefore, specifically such an effect limits the pulsed safe operation area (see Chapter 3) due to local burnout of the device structure [9] in the formed high-amplitude current filament. Implementation of the device-level negative feedback is one of the practical measures used to enable self-protection capability or improve pulsed SOA.

Since the pulsed current density of the semiconductor device in conductivity modulation mode can be up to 10 mA/μm, the physical limit of the operation regime is not necessarily always related to the semiconductor part of the device. The backend metallization and contacts should also be taken into account, especially in high-voltage devices. This mainly occurs due to electromigration limits because of high dissipated power and thermal heating of the backend region. For the given

process technology, the electromigration limits are usually a part of the process technology electrical design rules. The correlation between the electromigration rules and electromigration in the ESD pulse domain can be experimentally measured. To estimate backend limits in ESD pulse conditions the conventional electromigration limits can be multiplied by some process-specific factor. Usually, in 0.13–1 μm process technologies, with a good safety margin, this factor is above 40. A typical ESD device of standard 2 kV with peak current of 1.33 A usually requires $W =$ 50–1000 μm total width depending on the current density provided by the implemented elementary semiconductor structure. In this case, adequate backend metallization routing can be easily achieved following certain rules for clamp design (see Chapter 4).

A very important consideration should be emphasized at the end of this section. It is the fact that *local* structure-level current instability involves a necessity of implementation of device-level *local* current density suppression below a certain safe level. An intuitive attempt to use backend ballasting for this purpose is usually not successful if the current distribution in the finger of the device results in uncontrollable current density increase. Backend ballasting is used only to balance turn-on or current across multiple fingers [30].

Thus, the local positive feedback in the structure caused by conductivity modulation should be suppressed by some local negative feedback *on the device level*, although the spatial parameter for negative feedback can often be bigger. This measure is achieved by implementation of current ballasting regions inside the device that provide local current saturation effects. Total current limitation by the external circuit components is usually unsuccessful due to local current filamentation inside the device structure.

Practically, current density suppression in the ESD device is usually done so that all possible current filament states have limited amplitude, which ensures that heating during ESD pulsed domain (∼100 ns) is sufficiently below the critical level for structure damage.

However, this reliable operation does not absolutely guarantee that such a result is achieved in the event of much longer electrical overstress pulses or in dc conditions. Thus, ESD operation in the negative conductivity regime can be considered as an operation in the filament state with limited amplitude that in normal conditions would melt the device locally or result in other irreversible accelerated degradation effects. Physical limitations of the ESD device usually occur at a higher current level, when conductivity modulation of either the ballasting or contact regions results in formation of corresponding high-amplitude filaments followed by unlimited injection from the contacts.

2.1.4 Snapback Operation

As mentioned above, ESD device operation with negative differential resistance on a certain load resistance is usually called *snapback* in practical electrical engineering. In the case of snapback, the ESD device turns on or triggers on into

a high-conductivity or high-current state depending on the load characteristics in the circuit.

One of the most popular ways to measure snapback characteristics is the transmission line pulse (TLP) setup that usually provides a 50 Ω load [2].

In principle, there are no published standards that would require any TLP characteristics to be achieved by the integrated product. Historically, the capability of the HBM, MM, and other ESD testers provided mainly the results of pass–fail measurements, while the voltage waveforms of the ESD pulse operation were impossible to measure. Therefore, for device-level ESD design and debugging, a less expensive and more informative pulse method was required to compare different ESD devices and provide fast debugging of the IC's ESD failures.

The TLP is widely used currently, although a new state-of-the-art method that allows complete waveform measurements for an HBM pulse with a ∼1.5 kΩ load has been recently developed [32].

One of the major purposes of TLP evaluation is not only to confirm the maximum current level provided by the clamp but also to deliver insight on the most probable voltage waveform realized by the ESD device. This is achieved by extraction of several major figures of merit from the TLP characteristics.

From a physical point of view, the ESD device in snapback mode operates similar to a voltage-controlled switch with resistive load. The first pair of parameters, critical triggering voltage V_{T1} and triggering current I_{T1}, is used as figures of merit for turn-on of the device into snapback. These parameters are critical in providing both device turn-on within the so-called *ESD protection window* and guaranteeing that normal circuit pin operation will not result in the transient latch-up phenomenon.

The next important figure of merit parameter is the holding voltage V_H. This parameter depends on the load resistance and snapback voltage. Thus, for example, when using a 50 Ω TLP system, the measured holding voltage generally can be higher than the real minimum holding voltage for the example when measured using a waveform HBM system with a load resistance of 1.5 kΩ.

After turn-on into the high-current state, the device provides a certain on-state resistance due to the internal positive feedback and a saturation region that determines the voltage waveform in the ESD pulse domain at a high current level. Usually, this parameter is an important practical criterion. For standard package level specifications it can be defined, for example, as 1.33 A peak current, at HBM conditions of $V_{2\,kV}$.

Finally, at a certain point, physical limitation of the device operation results in irreversible changes to the device structure. This regime is usually named I_{T2} and V_{T2} (Fig. 2.1)

The parameters I_{T2} and V_{T2} in general cases cannot be seen from the pulsed *I–V* characteristic itself, because it may keep the same trend even after irreversible changes. A separate functional test is usually required to establish reversible operation in the device. In most TLP systems, this functional test is usually implemented by a simple leakage current measurement at a given voltage level with assumed parametric failure criteria, for example, a leakage deviation of one order of magnitude from the original level.

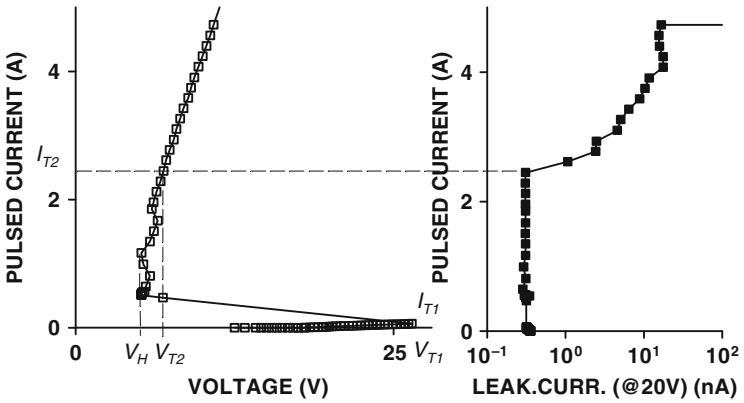


Fig. 2.1 Typical TLP snapback characteristics with major figures of merit indicated on the plot

In most TLP systems, the leakage voltage is measured between each pulse (Fig. 2.1). For example, in the case of data for the 20 V snapback device presented in Fig. 2.1, the left plot for pulsed *I–V* demonstrates no peculiarity associated with the irreversible failure that is already observed in the leakage current functional test plot at an I_{T2} current level of ~2.5 A.

TLP characteristics are very convenient for comparative analysis. They are widely used across this book both to represent the device parameters, pulsed SOA, and for debagging of the analog circuit product pin characteristics.

A new methodology combines the pass–fail test with waveform capture [33]. It is presented below as well, as an important future ESD design tool that enables a much faster time-to-market.

2.1.5 Notes to the Methodology of Material Presentation in This Chapter

The elementary semiconductor structure presented below can be used to control the pulsed characteristics of the developed device and clamp in a wide range to achieve a desirable voltage waveform at ESD pulse conditions.

The purpose of the following section is to provide a condensed, thorough background for understanding of conductivity modulation in the semiconductor structure. This knowledge, combined with a certain level of physical understanding, is directly relevant to both ESD devices and clamp design.

This goal can be achieved in several different ways. In most books published in this field, authors prefer presenting analytical descriptions of the phenomena or extensive experimental data.

We believe that in most practical cases, the analytical approach requires a significant trade-off between the complexity of real objects and simplifications required for the analytical description itself. Although such an approach is fit for teaching on

a physical level, there is little practical use of such material for a wide audience of engineers and designers, due to the diminished practical relevance of the simplified structure parameters and the physical models of semiconductor material and specific devices. Thus, this approach is usually applicable within a rather narrow range of parameters and provides limited accuracy of conclusions that greatly depend on the expertise and experience of the theoretician.

Without questioning the usefulness of analytical approach in general, we believe that a much broader pool of readers with more practical needs will appreciate an alternative approach specially "engineered" for this book, which takes advantage of modern physical processes and device simulation tools.

This alternative approach combines empirical description with rather accurate numerical simulation results.

The advantage of such an approach lies in the ability to avoid excessive simplification on the semiconductor level in simulation, while still presenting the data using accurate representations of the drift-diffusion model for carrier transport equations and semiconductor material parameters for silicon. At the same time, a complete visualization of the intrinsic structure state is fully possible.

Namely, this simulation approach is followed consistently across the book chapters in each of the suggested methodology levels.

However, there are some important limitations involved. In this chapter, when describing the elementary conductivity modulation structures, we will limit ourselves within the following constraints:

(i) Silicon semiconductor material parameters
(ii) Isothermal cases specific to ESD pulse domain
(iii) Selecting the doping profiles and structure dimensions close to parameters of the integrated structures in 0.5 μm process, as this process technology is most widely used today for analog design
(iv) Staying within practical limits of the current densities in conditions where conductivity modulation will be limited below 10 mA/μm^2 level

In spite of these assumptions, both the description and the simulation results could be easily expanded for broader cases.

How successful this approach will prove to be will be known through the future feedback of those who read this book. Meanwhile, possible gaps in understanding can be covered by the recommended books in the field where the analytical approach to learning is presented in greater detail [1–9, 14, 26].

2.2 Avalanche Breakdown in Reverse-Biased p–n Structure

In ESD devices, avalanche breakdown of the blocking junction is a physical phenomenon that is typically used to initiate the self-triggering of a voltage referenced in snapback ESD devices. Avalanche diodes can be used directly to enable voltage

limitation by the avalanche current generation. The alternative way to initiate snapback is to implement a triggering circuit that is often based on avalanche breakdown voltage reference as well.

In the case of snapback devices, conductivity modulation and current instability follow the avalanche breakdown mode. The same phenomenon is responsible for the physical limitation of pulsed SOA of standard semiconductor devices supported in the given process technology in the ESD time domain.

In this section, avalanche breakdown is first discussed using simplified analytical models. Then, a simple quasi-1D numerical simulation analysis using a 2D simulation tool is presented in order to demonstrate the internal effects of the structures.

This is the only place in the entire book where the analytical approach is compared to the methodology of combining the phenomenological approach with numerical simulation, as implemented in this text.

For ESD devices, the reverse-biased p–n junction presents a direct interest in several aspects. First of all, this structure forms a blocking junction that provides a corresponding high-voltage tolerance with low leakage of the self-triggered ESD device. The positive feedback and current instability inside the ESD device can usually be initiated at a relatively low level of the avalanche current. Therefore, the contribution of an additional voltage drop in avalanche mode is rather low. This automatically provides a correlation between the avalanche breakdown voltage and the snapback triggering voltage in a more complex device.

Typically, the high-current operation of the avalanche diode device can produce current levels of 0.1 mA/μm. This makes the avalanche diode directly useful not only as a source of reference voltage and current, but as a clamp. In principle, the total width for a packaged-level spec for lateral avalanche diodes of the clamp is rather large: ~10,000 μm. Therefore, a more practically efficient way to implement the ESD protection circuit is the use of avalanche diodes in the second stage of the two-stage ESD protection network. In this case, the avalanche diode is typically designed to provide current of 1–50 mA that can be achieved at a relatively small size of laterally or vertically self-aligned or non-self-aligned devices (see Chapter 4).

2.2.1 Analytical Description of the Avalanche Breakdown Phenomenon

This section provides an analytical description of avalanche breakdown.

Generation of the electron–hole pairs due to impact ionization in the semiconductor structure becomes evident at an electrical field of $E > 10^4$ V/cm. At some critical level of electric field, this process has an avalanche nature. The impact ionization coefficients α_n for electrons and α_p for holes are used for analysis of the avalanche breakdown process. The coefficient denotes the number of ions generated by the charged particle on the pathway unit. The coefficient is a strong function of the electric field. It is given by exponential dependencies [14]:

$$\alpha_n, \alpha_p \approx \exp\left(-\frac{E}{E_0}\right)$$

In the case of 1D asymmetric reverse-biased p^+–n junction, the coefficient dependency may be written in the following way, without taking into account the diffusion and recombination in the balanced equation:

$$-\frac{dj_n}{dx} = \frac{dj_p}{dx} = \alpha_n j_n + \alpha_p j_p, \tag{1.1}$$

where $j_n = qnv_n$, $j_p = qnv_p$ are the electron and hole current densities, v_n, v_p are the electron and hole drift saturation velocities that are independent from E in a strong field: $v_n \approx v_p = 10^7$ cm/s. The boundary conditions for the p^+–n junction are given by

$$j_n(x = 0) = 0; \quad j_p(x = W) = j_s, \tag{1.1a}$$

where W is the width of the space charge region (SCR) of the p^+–n junction and j_S is the saturation current density. The current discontinuity equation is added to (1.1):

$$j = j_n + j_p = \text{const.} \tag{1.2}$$

From the solution of (1.1) and (1.2) at $\alpha_n = \alpha_p = g$, the current density is

$$j = \frac{j_s}{1 - \int_0^W g\,dx}. \tag{1.3}$$

The ratio of the total current density to the saturation current density j/j_S is denoted as the multiplication coefficient M:

$$M = \frac{1}{1 - \int_0^W g\,dx}. \tag{1.4}$$

For the conditions of high avalanche breakdown $j \to \infty$, $M \to \infty$, $\int_0^W g\,dx \to 1$.

Usually, as a criterion of the avalanche breakdown, a certain given integral value $\delta << 1$ is accepted:

$$\int_0^W g\,dx = \delta \tag{1.5}$$

Since the total current density $j(E)$ is a sharp function of E, most of the avalanche current in the p^+–n junction is generated in rather narrow layer $\Delta << W$. Therefore, the breakdown condition (1.5) can be approximated by

$$g\,(E_{BR})\,\Delta = \delta. \tag{1.6}$$

Thus, the defined electric field E_{BR} can be adopted as the avalanche breakdown value. This value is changed slightly at Δ and δ variation.

For analytical estimation of the *I–V* characteristic in the avalanche breakdown mode, the following assumptions are used in addition to the assumptions above:

(i) Uniform distribution of the breakdown across the uniform p–n junction area with the lateral dimension of the p–n junction significantly higher than its thickness and thus the 1D problem is automatically justified.
(ii) The drift velocity v does not depend upon the electric field starting from the levels of electric field above $E > 10^4$ V/cm.
(iii) $E(x)$ is a smooth function of coordinate, with the simple function of ionization coefficient $g \approx E^m$.

Under these assumptions, the basic equation for *I–V* dependence is given by the Poisson equation for the 1D case:

$$\frac{dE}{dx} = -\frac{\rho}{\varepsilon}, \tag{1.7}$$

where ρ is the charge density and ε is the dielectric constant of semiconductor material.

In case of developed avalanche breakdown ($j_S > j$), the space charge density in the n-region is given by

$$\rho = qN_D - \frac{j}{v}, \tag{1.8}$$

where N_D is the donor concentration, j is the electron current density in the n-region, and the avalanche generation is concentrated in the multiplication layer $\Delta << W$.

From [14], the voltage on the p–n structure U is presented by the expression

$$U = U_{BR}\left(1 - \frac{j}{j_N}\right)^{-1}, \tag{1.9}$$

where $U_{BR} = \varepsilon E_{BR}^2 / 2qN_D$ and $j_N = qN_D v$ is the critical current density at total space charge density equal to zero.

At $j \rightarrow j_N$, the *I–V* dependence $I(U)$ of the p–n junction aspires to saturation. At this condition, $dI/dU > 0$, i.e., the p–n junction has positive differential conductivity.

2.2.2 Numerical Analysis of the Avalanche Breakdown in the p^+–p–n^+ Structure

In the above section, an example of the analytical approach is demonstrated to illustrate the most basic phenomenon. It provides an in-depth understanding of

the physical nature of the phenomenon that is very useful in practice. However, the application of the analytical approach to integrated ESD device engineering is not straightforward. For example, the application of the analytical approach to practical design of the ESD avalanche diode in the given integrated process technology with complex profiles, multiple interface regions, and reduced surface field (RESURF) effects would require numerous simplifications and assumptions. Thus, this approach would hardly guarantee appropriate accuracy.

The practical alternative to this method is a numerical simulation of devices using finite element models (FEM). Numerical simulation is actually based on the analytical model for carrier transport of semiconductor devices.

Numerical FEM methods can be used to simulate devices with arbitrary geometry and doping profiles in 2D or 3D and can be calibrated and solved with an accuracy that can hardly be achieved with the analytical approach, while also providing fast simulation speed.

This approach does not imply that numerical simulation requires no expertise. Previously, exclusively TCAD engineers ran the industrial TCAD software, which involved writing substantial amount of code and extensive knowledge of the tool-specific software limitations and peculiarities. Today, with a new interactive tool DECIMM^{TM} from Angstrom Design Automation [31], the learning overhead is significantly reduced. The DECIMM^{TM} tool is designed to be user-friendly and usable by any electrical engineer, both device and circuit designers, and to provide results that can be directly used in engineering work. This in particular is exploited across this book by the complementary examples referred to at the end of each chapter.

In the case of real device, the p–n structure in integrated process technology usually has a more complex architecture. At least two heavily doped contact regions are required: p^+ and n^+ regions. Then, at least one lightly doped region is required to tolerate practical levels of operation voltage not limited by the tunneling effect.

An example of simulation structure used to analyze avalanche breakdown phenomena in the silicon p–n junction is presented in Fig. 2.2. The solution for different doping levels in the p-drift region (or p-base) (Fig. 2.3) can be obtained within seconds providing an accurate solution for initial avalanche breakdown voltage at low current level, and the saturation *I*–*V* dependence. Moreover, the solutions for all internal characteristics, for example, electric field and carrier densities (Fig. 2.4a–c), can be easily plotted from the simulator output for different current levels, helping to reveal the internal processes in the device.

The particular device in this example has an $L = 3$ μm space between the *anode* (positively biased) and *cathode* (negatively biased) contacts with a $W = 1$ μm width.

Further in this chapter, we will deviate from the commonly used approach of calling the electrodes of the diode structure that are subjected to the positive and negative bias *anode* and *cathode*, respectively. This is done to avoid misunderstanding and confusion of avalanche diodes and other diode structures, for example, with forward-biased diodes.

Current density is normalized for a 1 μm structure depth. Three uniform doping profiles are defined in the silicon region of the structure: p^+ region with acceptor concentration $N_{Ap+}=10^{20}$ cm^{-3} defined from $y = 0$ to position $y = 0.5$ μm, n^+-region with donor concentration $N_{Dn+}=10^{20}$ cm^{-3} from $y = 1.5$ μm up to the end of

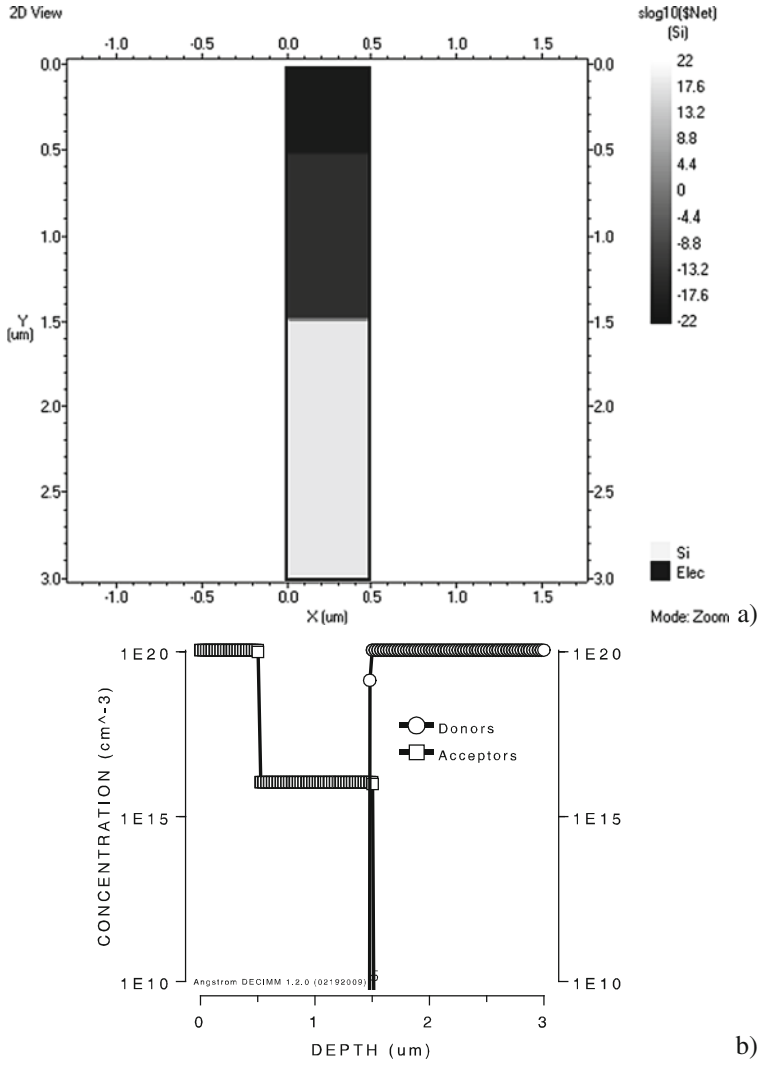


Fig. 2.2 Diode p^+–p–n^+ structure with additional drift regions (**a**) and 1D cross section through the center of the structure showing detailed doping profile (**b**)

the device. Uniform doping across the whole structure p^- with N_{Ap} is variable and defines the drift or a base region of the structure with a length of $L_P = 1$.

In agreement with the analytical approach in the simulation results (Fig. 2.3), a decrease of the breakdown voltage is observed with an increase of the drift region acceptor doping. The avalanche breakdown current is generated upon the predicted exponential dependence up to a certain current level, when current saturation effect in the drift region becomes dominant. With further current increase,

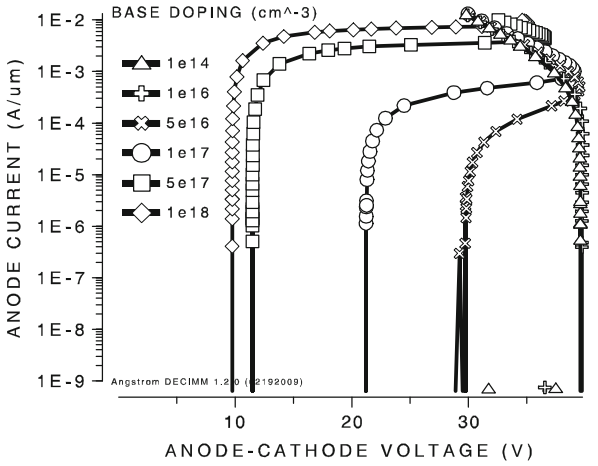


Fig. 2.3 Calculated *I–V* characteristics of the structure in Fig. 2.2 of different N_{Ap} doping values

a current instability provides for the appearance of the S-shaped *I–V* portion with negative differential resistance (Fig. 2.3). According to the analysis of the internal parameters distribution, it can be easily concluded that the observed current instability in this avalanche diode example is the result of p-base region conductivity

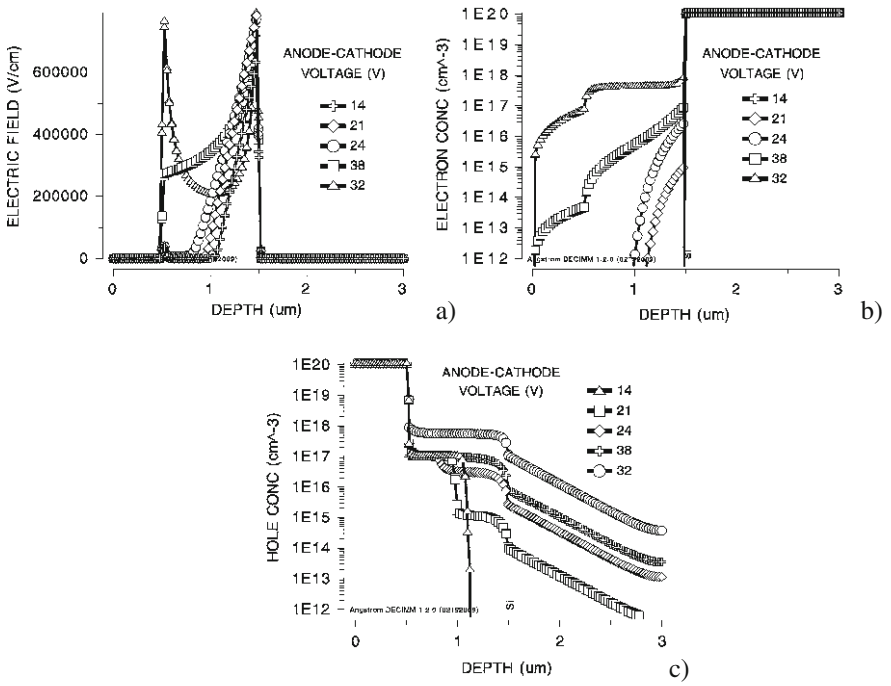


Fig. 2.4 Calculated distributions of the electric field (**a**), electrons (**b**), and holes (**c**) for the structure with $N_{Ap}=10^{17}$ cm^{-3} at various bias and current levels

modulation after the level of injected carriers exceeded the level of acceptor doping.

Thus, the process in this structure can be easily analyzed in greater detail by comparing the magnitude of the electric field and the carrier density (Fig. 2.4a–c). Cutline profiles are taken for the device with $N_{Ap}=10^{17}$ cm^{-3} in the middle of the structure in the Y-direction for different bias and current levels. At a low current level of $\sim 10^{-10}$ A (bias $\sim$14 V), the electric field profile forms a triangular shape at the p–n$^+$ junction peaking at $Y = 1.5$ μm.

As the current increases up to $\sim 10^{-5}$ A, it is supported by the exponential dependence of the avalanche multiplication peaking at $Y = 1.5$ μm. After the current saturation at voltage level of $\sim$24 V, the electric field increase begins in the base region, thus providing a higher integral voltage drop. The carriers injected into the base region provide partial space charge neutralization, which results in the increase of the electric field magnitude in the p-base. This occurs up to the formation of the trapezoidal electric field distribution in the p-base at a voltage level of $\sim$38 V (Fig. 2.4a).

The high electric field provides carrier multiplication across the whole p-base region, while the injected carrier density is above the $N_{Ap}=10^{17}$ cm^{-3} level. This effect produces a flow of avalanche-generated electrons and holes drifting in the strong electric field in opposite directions across the p-base. At high current level, the space charge of the carriers is compensated, providing the electric field distribution with two maxima at 32 V (Fig. 2.4a).

Thus, at the high current density typical for ESD operation, the conductivity of the p-region is modulated and the structure provides negative differential resistance. As it will be demonstrated in Section 2.8, the current filamentation effect further results in high-amplitude current filament formation. Unless additional contact saturation region is provided in the structure, the density of the carriers injected into the filament will reach the doping level in the contact regions, resulting in tremendous locally generated power. These phenomena, in most practical cases, are responsible for local burnout of the reverse-biased p–n diodes.

2.3 Double-Avalanche–Injection in p–i–n Structures

2.3.1 An Analytical Description of the Effect

At first, operation of this structure in reverse bias is somewhat similar to the operation of the reverse-biased p–n junction in avalanche breakdown mode. However, the presence of the i-region creates an interesting positive feedback effect. This effect is essentially similar to the p-base modulation in the avalanche diode discussed in the previous section, but observed at a much lower current level.

An analytical description of the structure with simplified parameters is presented below, followed by more detailed numerical simulation results with exact model parameters.

On the empirical level, the conductivity modulation of the i-region in the p–i–n structure is rather easy to understand.

At reverse bias, the main part of the potential difference U is supplied to the i-region of the p–i–n structure. At $U \approx U_{\mathrm{BR}}$, the electric field in the i-region is constant and impact ionization begins uniformly through the i-region.

The generated electrons in the i-region are accumulated near the n-region and the generated holes near p-region. The result of this distribution is the increase of the electric field near the junctions and decrease in the middle of the structure due to space charge mutual neutralization.

The negative differential resistance of the p–i–n structure is easy to prove using the analytical approach; the p–i–n diode *I–V* characteristic can be determined by the following equations. The electric field in the i-region at $I = \delta I \neq 0$ is given by

$$E = E_{\mathrm{BR}} + \delta E\,(x, \delta I)\,. \tag{1.10}$$

For the initial part of the characteristic at $\delta E/E_{\mathrm{BR}} \ll 1$, the changing electric field δE is limited by the breakdown condition:

$$\int_0^W g\,(E)\,\mathrm{d}x = \text{const.} \tag{1.11}$$

By substitution of (1.10) into (1.11), expanding $g(E)$ and limiting by the first three terms

$$\delta U = \int_0^W \delta E \mathrm{d}x = -\frac{1}{2}\frac{g'_{\mathrm{BR}}}{g''_{\mathrm{BR}}}\int_0^W (\delta E)^2\,\mathrm{d}x. \tag{1.12}$$

Since $g'_{\mathrm{BR}} \equiv \left(\mathrm{d}g/\mathrm{d}E\right)_{E_{\mathrm{BR}}}$ and $g''_{\mathrm{BR}} \equiv \left(\mathrm{d}^2 g/\mathrm{d}E^2\right)_{E_{\mathrm{BR}}}$ are positive, the increment of the voltage drop δU is negative. This simple fact means that at least the internal differential resistance of the p–i–n diode at avalanche breakdown is negative.

2.3.2 Numerical Analysis for the p–i–n Diode Structure

As it was mentioned before, the numerical solution for the p–i–n structure is similar to the high-current operation of the reverse-biased p^+–n–n^+ or p^+–p–n^+ structure at the current level that corresponds to the density of injected carriers above the base (drift) region doping level. However, at low current levels, the electric field distribution is different and the saturation region is not present. Essentially, from the conductivity modulation point of view, the p–i–n structure is the particular case of p^+–n–n^+ or p^+–p–n^+ structure where the base doping level approaches zero.

A simulation study of the p^+–i–n^+ is presented in Figs. 2.5 and 2.6. This structure is geometrically identical to that described in Section 2.2.2 with N_{Ap}

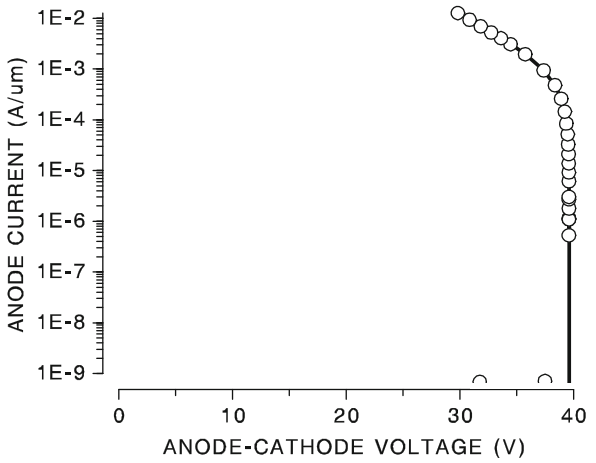


Fig. 2.5 Numerical analysis of the conductivity modulation in the p–i–n structure. *I–V* characteristics

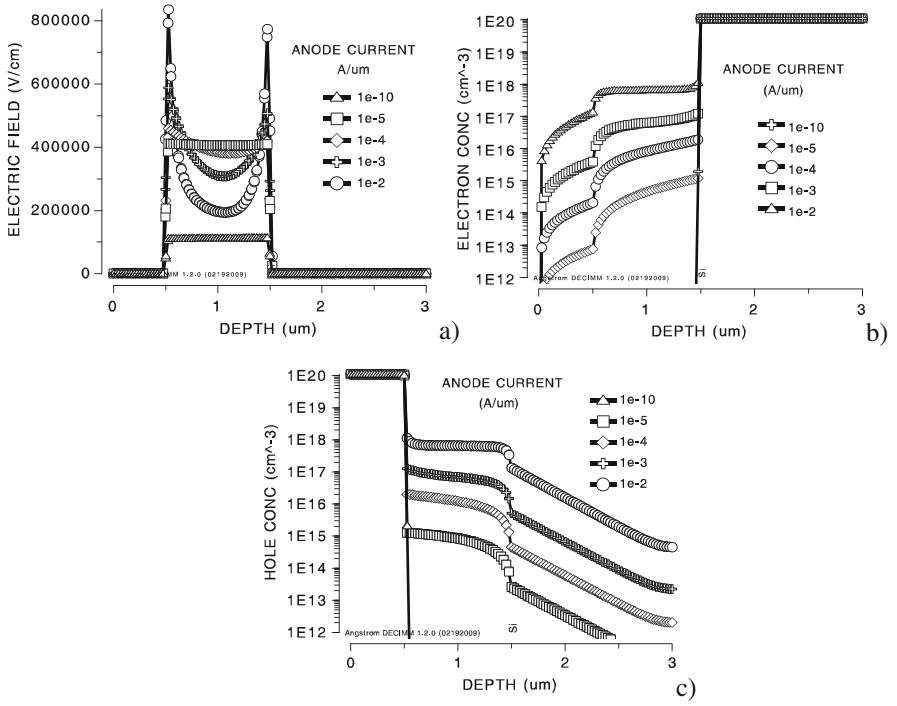


Fig. 2.6 Numerical analysis of the conductivity modulation in the p–i–n structure; distributions for the electric field magnitude (**a**), electron (**b**) and hole density (**c**) at selected bias and current levels

set to 10^{14} which is equivalent to intrinsic Si from the practical point of view. Simulated *I–V* curve is shown in Fig. 2.5. The distribution of electric field magnitude, electrons, and holes in the middle of the structure in the *Y*-direction is summarized in Fig. 2.6a–c. With the voltage increase, the original electric field distribution in the structure's i-region is uniform. At the breakdown voltage level, the energy of the carriers is sufficient to cause the impact ionization effect that is initiated across the whole i-region. This forms a flow of electrons and holes in the opposite direction. At some critical level, the quasi-neutral area in the middle of the i-region provides for a decrease of the electric field due to partial space charge compensation of the injected carriers. At the same time, the total avalanche current increase is supported by a slight increase of the electric field magnitude at the n^+–i and p^+–i junctions (Fig. 2.6, plot for current 10^{-4} A/μm). With current increase due to exponential dependence of the multiplication coefficients upon the electric field, the total voltage drop on the multiplication regions is less than the total voltage drop in the partially quasi-neutral middle of the i-region (Fig. 2.6, plot for current 10^{-3} A/μm). As a result, these internal processes provide for negative differential resistance of the entire device (Fig. 2.6, plot for current 10^{-2} A/μm).

2.4 Avalanche–Injection in Si n^+–n–n^+ Diode Structure

The structure of the n^+–n–n^+ diode is the best representation of the saturation resistor structure that is often used in ESD circuits for two stage ESD protection networks (Chapter 4).

The structure of the n^+–n–n^+ diode (Fig. 2.7a) is helpful for modeling avalanche–injection in an n^+–p–n–n^+ bipolar transistor for active regime operation (the positive n^+-contact is a collector equivalent). The physical processes in this Si structure are very common for other materials too. For example, the n^+–i–n^+ parasitic structure is formed by source, i-buffer, and drain regions and is responsible for current instability in GaAs MESFETs [18], similar to how the n^+–p–n^+ parasitic structure is responsible for the same effect in discrete Si NMOSFET devices.

As for integrated electronic components, the structure represents the drift region in the high-voltage devices, drain extended MOS in CMOS processes, N-channel lateral double-diffusion MOS (NLDMOS) devices fabricated in conventional BCD processes or a lateral insulated gate bipolar transistor (LIGBT), and many other high-voltage devices including avalanche diodes. Thus, understanding the conductivity modulation in such devices is rather important.

Similar to the previous sections, the n^+–n–n^+ structure is first discussed within an analytical approach. Then, the numerical simulation example is provided to demonstrate operation in high injection conditions for a case close to typical analog BiCMOS and BCD processes.

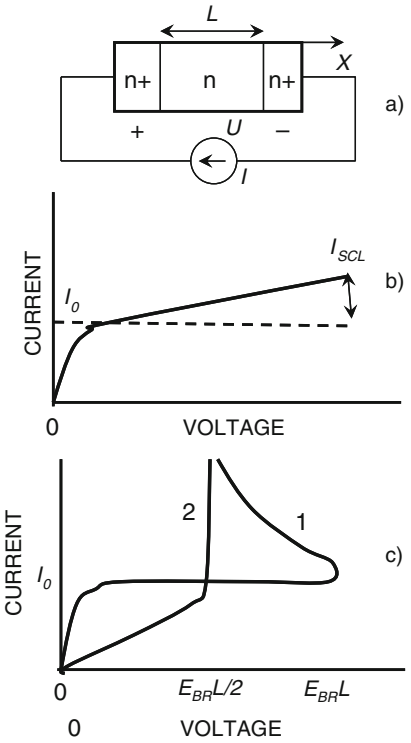


Fig. 2.7 Structure of the n^+–n–n^+ diode (**a**) and typical *I–V* characteristic (**b**) for the boundary cases of SCN type "1" and SCL type "2" (**c**)

2.4.1 Analytical Approach

For an n^+–n–n^+ diode at a fixed current I, Ohm's law is true only at rather small current levels. In this case, the resistance of n-region R_0 is equal to

$$R_0 = \frac{L}{q\mu_n N_D A}, \tag{1.13}$$

where N_D is the donor concentration and A is the total diode area $\left(\sqrt{A} \gg L\right)$. With further current increase, the electric field $E = U/L$ increases up to the level $E > 10^3$ V/cm, where the drift velocity is saturated at the level $v \approx 10^7$ cm/s.

The density of free electrons in the n-region is equal to $n = j/v$. At the dielectric relaxation time in the n-region that is less than electron drift time, $n = N_D$ is true and the current density is saturated at level $j_0 = qvN_D$.

In the opposite case when $n > N_D$ and the total space charge is negative $\rho = qN_D - j/v$, the *I–V* dependence of the n^+–n–n^+ diode at $j > j_0$ is not saturated and has a slope equal to $dI/dU = 2\varepsilon vA/L^2$.

A typical experimental *I–V* characteristic of the n^+–n–n^+ diode is presented in Fig. 2.7b. For adequately high U, the diode current I is equal to the sum $I = j_0 A +$

I_{SCL} , where I_{SCL} is the space charge limited current component (SCL). The ratio between I_0 and I_{SCL} depends upon U, N_D, and L.

For theoretical analysis, an understanding of two boundary cases is rather helpful: space charge neutralized (SCN) diode and SCL diode (Fig. 2.7).

In the SCN diode $\rho = 0$, the electric field is distributed uniformly: $E = U/L \approx const$, I–V characteristic has a saturation region at $I = I_0$.

In the SCL diode $\rho < 0$, the electric field is increased linearly in the anode contact direction: $E_{MAX} = E(0) = 2U/L$. Near the cathode contact, $E(L) \approx 0$ and the electron injection in the n-region is rather limited.

Since both the doping concentration and the length of the diode play a major role, a relative boundary between the SCN and SCL diodes for Si devices is $N_D L = 6 \times 10^{11}$ cm^{-2} [35]. At $N_D L \ll 6 \times 10^{11}$ cm^{-2} the diode is SCL; at $N_D L \gg 6 \times 10^{11}$ cm^{-2} the diode is SCN. However, this statement is true with the assumption that the diffusion length of major carriers is bigger than the device length L, $L_D \sim L$.

At a small diffusion length or a long n-region, for example, due to a high recombination rate, the SCL case is dominant. This situation is also realized in GaAs n^+–i–n^+ diodes.

Using the avalanche breakdown criterion (1.6), the avalanche multiplication in the SCN diode begins at $U \approx E_{BR}L$. Electrons are accumulated near the anode electrode; the field is increased in this region and shifted in the multiplication region near the n^+–n boundary. The holes drift to the cathode, thus being injected into the n-region, namely this phenomenon resembles the notion of "avalanche–injection."

This uncompensated hole space charge causes injection of electrons from the n–n^+ junction. At $I >> I_0$, a multiplication region is formed at the anode region and the maximum field slightly exceeds E_{BR} due to a sharp dependence $g(E)$. The negative space charge $\rho = N_D - n + p$ in the multiplication region provides electric field decrease to $E(L) \approx 0$. This SCL diode regime is realized at $U = (E_{BR}L)/2$.

With the current increase above $I > I_0$, the SCN diode voltage is decreased twofold (Fig. 2.7c) and an S-shaped I–V characteristic is observed in the critical regime $I = qN_D vA$; $U \approx E_{BR}L$. Obviously, avalanche multiplication for the SCL-diode begins once $U = E_{BR}L/2$ and a negative differential conductivity region is not presented in the I–V characteristic (Fig. 2.7c).

In the n^+–n–n^+ structure in the avalanche–injection condition $M = j/j_n$ [9]; therefore, in high injection $j \to \infty$, $j_n \to j_p$, and $M \to 2$. This fact namely provides the difference between the avalanche breakdown process in reverse-biased p–n junctions, where $M \to \infty$, and the avalanche–injection conductivity modulation that occur at rather low $M \to 2$.

From the experimental study [35], for samples with various N_D and L, it follows that the I–V characteristic of real n^+–n–n^+ diodes has an intermediate shape between the boundary cases of the SCN and SCL diodes.

It should be emphasized again that the above conclusions are true for cases where diffusion length $L_D > L$. For a diode structure with $L_D < L$, the S-shaped I–V characteristic is realized independently of the type and doping level. For example, in n^+–i–n^+ structures, the deep S-shaped I–V characteristic is observed both in experiments and in numerical simulation studies [9].

2.4.2 Simulation Analysis

A significant expansion of understanding can be achieved at the simulation level. A summary of the isothermal quasi-static numerical simulation analysis for the n^+–n–n^+ structure case is presented using the corresponding diode structure (Fig. 2.8). This structure is geometrically identical to that described in Section 2.2.

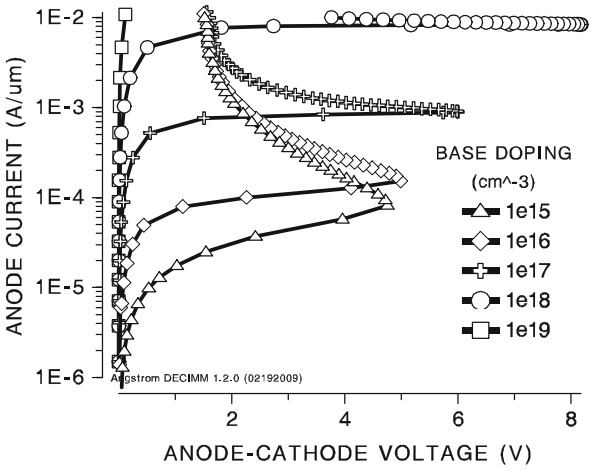


Fig. 2.8 Isothermal *I–V* characteristics of n^+–n–n^+ diode structure calculated for different n-region doping levels

The device has a similar $L = 3$ μm space between the *anode* (positively biased) and *cathode* (negatively biased) contacts with a $W = 1$ μm width. Current density is normalized for the 1 μm structure depth. Three uniform doping profiles are defined in the silicon region of the structure: the cathode n^+ region with acceptor concentration $N_{Dn+}=10^{20}$ cm^{-3} defined from $y = 0$ up to position $y = 0.5$ μm and the anode n^+-region with donor concentration $N_{Dn+}=10^{20}$ cm^{-3} from $y = 1.5$ μm to the end of the device. Uniform doping is defined across the whole structure n^-, with N_{Dn} as a variable parameter that defines the doping level in the n-drift (or the n-base) region of the structure with fixed length $L_P = 1$ μm (Fig. 2.8).

The simulated quasi-static isothermal *I–V* characteristics of the device with different uniform doping levels are presented in Fig. 2.8, practically demonstrating that the SCN structure is formed for the given structure parameters.

According to the analytical solution in the previous section, the device indeed provides a linear resistance, followed by the formation of the saturation region as a function of the n-region doping level.

After reaching the high injection current level, the conductivity modulation in this structure is observed. This phenomenon is similar to the corresponding physical effect in n–p–n diodes discussed in the following section. It is related to the positive feedback between avalanche multiplication at the anode junction and injection from the cathode junction. This will be demonstrated by comparative analysis of the electric field and carrier density distributions at different current densities.

2.5 Conductivity Modulation Instability in n–p–n Diode Structures

Isothermal current instability (or the current mode secondary breakdown) was historically first observed in power epitaxial n^+–p–n–n^+ structures. The first studies demonstrated that the thickness L and the doping level N_D in the n-region provide a considerable impact on the condition of isothermal instability. Particularly, it was determined in [34] that in structures with a thick epitaxial layer and in transistor structures without the n-layer, S-shaped *I–V* characteristics are not formed in a rather wide current range. At the same time, in other BJT structures, isothermal current instability has been observed at a relatively small collector current, followed by catastrophic failure.

Different types of breakdown S-shaped *I–V* characteristics were measured using a high-speed experimental technique. The particular type of technique used depends on the structure, operation regime, and transistor type of the device.

From ESD design point of view, n–p–n devices are usually considered both as parasitic structures formed in the integrated design and as components that provide the conductivity modulation effect for the ESD device itself. However, at a submicron dimension, the diode n–p–n structure provides a rather high-temperature-dependent leakage. Therefore, instead of a diode with a floating base, a triode n–p–n structure is used, with the p-base held at the same initial potential as the n-emitter. The triode structure or simply the bipolar junction transistor structure is discussed in Section 2.6.

2.5.1 Conductivity Modulation in a Floating Base Region: Diode Operation Mode

The n–p–n structure essentially represents the case of a bipolar junction transistor in common emitter operation with an open base circuit.

For simplification of the following discussion, the 1D n^+–p–n–n^+ diode structure will be described under avalanche breakdown conditions.

2.5.1.1 The Case of Floating Base Breakdown (BVCEO) $I_B = 0$

The operation regime of transistors with a floating base is roughly similar to the n^+–n–n^+ diode described in the previous chapter. However, this transistor operation has several significant peculiarities.

Two versions of the transistor structure with different doping levels in the n-collector region present a particular interest: A structure with $N_D L \approx 6 \times 10^{11}$ cm^{-2} ($L = 10\ \mu m$) and C structure with $N_D L = 6 \times 10^{12}$ $cm^{-2} >> N_D L = 6 \times 10^{11}$ cm^{-2} ($L = 9\ \mu m$).

The C-type n^+–n–n^+ diode corresponds to the SCN type. The A-type diode is the intermediate case between the SCL type and the SCN type.

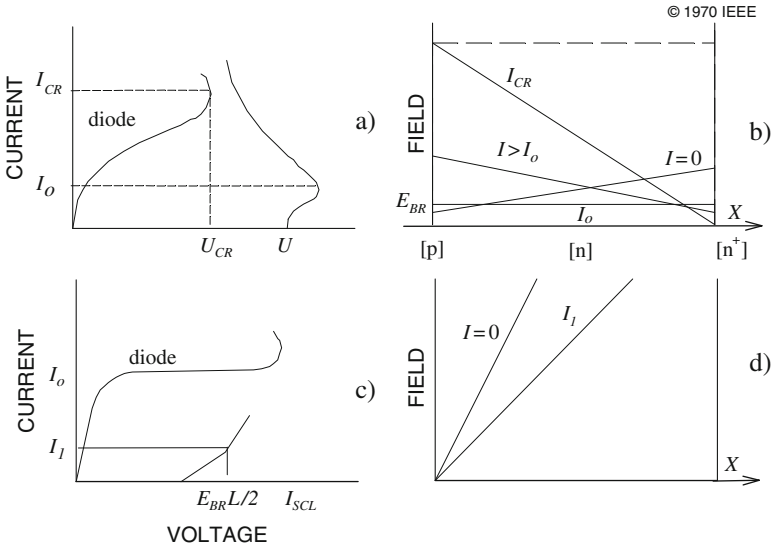


Fig. 2.9 Calculated *I–V* characteristics and electric field distributions for the structures with $N_D L = 6 \times 10^{11}$ cm^{-2} (**a**, **b**) and $N_D L = 6 \times 10^{12}$ cm^{-2} (**c**, **d**) at different current levels [35]

The shape of the *I–V* characteristics of the transistor structures and the electric field distributions in the n-region are presented in Fig. 2.9, in comparison with the *I–V* characteristics of the corresponding n$^+$–n–n$^+$ diode structures.

In operation regime with base current $I_B = 0$, the collector current is equal to the emitter current and proportional to $(1 - \alpha M)^{-1}$. The increase in current begins under conditions $U_{CE} \to U_\alpha$, $(\alpha M \to 1)$, i.e., when the electric field is approaching a certain maximum value $E_{MAX} \approx E_{BR}$ on the p–n junction (Fig. 2.9b). Since the value E_{MAX} is limited by the condition $\alpha M = 1$, below E_{BR} the initial current increase is observed at an increase of the collector–emitter voltage starting at level $U_{CE} \approx \int_0^L E\,\mathrm{d}x$.

In the "A" structure, at $j = j_0 = qN_D v$, the electric field levels off in the n-collector region as a result of the electron space charge. Due to a lower n-collector doping level at $j > j_0$, the avalanche multiplication region is formed on the n–n$^+$ junction, followed by a voltage decrease. The *I–V* characteristics of the A-type transistor structure have a negative slope and aspire to the *I–V* characteristics of the corresponding n$^+$–n–n$^+$ diode structure. At significant current density $I > I_{CR}$, both the transistor and the diode have similar characteristics.

For the more heavily doped structure "C," the I_0 value is so high that breakdown on the p–n junction is initiated when the space charge region (SCR) is confined within the n-collector region (Fig. 2.9d). In this case, the *I–V* characteristic has a positive slope that increases at $U \to E_{BR}L/2$, when SCR attains the n–n$^+$ junction. Therefore, for the C-type structure with a higher n-collector doping level, the differential resistance remains positive $\mathrm{d}I/\mathrm{d}U > 0$ up to the current level I_0, which

exceeds the corresponding level in the A-type structure approximately by one order of magnitude.

2.5.1.2 Numerical Solution for $I_B = 0$ Case

A summary of the simulation analysis is presented in Fig. 2.10. It reveals similar regularities to the analytical approach above and provides a better accuracy for the 1 μm p-base region. The quasi-neutral region formation at the emitter junction with a corresponding reduction of the electric field can be observed through comparison of the electric field and carrier density profiles at different current levels (Fig. 2.10c). However, a more accurate model demonstrates that the SCL regime is

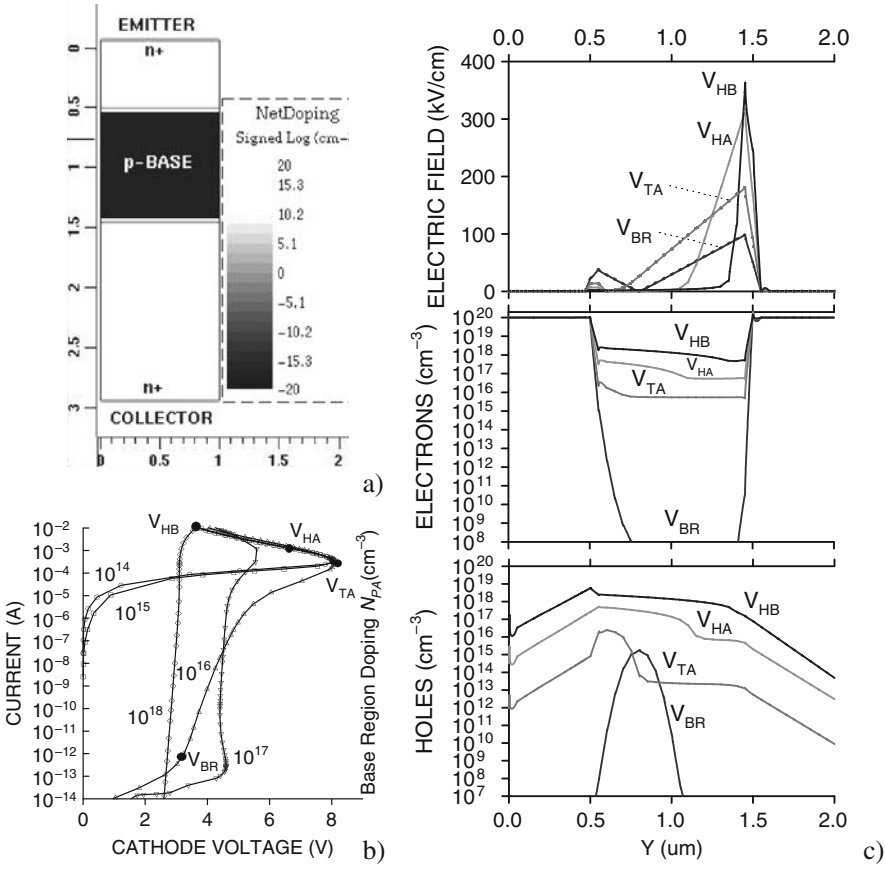


Fig. 2.10 n–p–n diode structure (**a**), simulated *I–V* characteristics for different base doping levels (**b**), and distributions of the electric field magnitude, electron and hole density at representative current levels (**c**) for the structure with p-base doping level 10^{16} cm^{-3}

actually observed above the level $N_A L = 10^{17} \times 10^{-4} = 10^{13}$, thus bringing the accuracy of the analytical estimation within one order of magnitude.

2.6 Conductivity Modulation in the Triode n–p–n Structure

The most important applications of the n–p–n structures in submicron semiconductor processes involve the triode operation regime. To analyze such an operation mode in numerical simulation, a third contact region should be provided, thus essentially making the device structure equivalent to the bipolar junction transistor (BJT) structure.

2.6.1 The Case of Grounded Base Breakdown Operation $U_{EB} = 0$ (BVCES)

In case of shorted emitter and base terminals, the initial location of avalanche breakdown is observed at the collector p–n junction. In this case, avalanche breakdown is not accompanied by injection from the closed emitter junction, due to the contact potential difference. Therefore, the collector current is provided by multiplication of the saturation current I_{C0} $I_C \sim MI_{C0}$. The breakdown voltage is $U_{CEBR} > U_\alpha$ and is practically equal to the collector–base breakdown voltage U_{CBBR}.

The avalanche-generated holes flow out of the base in negative base current regime. This hole current results in an additional voltage drop across the p-base, which results in an increase of the base potential at the emitter junction on the corresponding value $r_B I_C$, where r_B is the corresponding internal base resistance. The r_B can be extracted for the simulation or measurement data. Thus, when the voltage level $r_B I_C$ is equal to or above the potential of the emitter junction opening (approximately 0.7 V at room temperature conditions), injection of the electrons begins from the emitter, followed by a positive feedback mechanism. Thus, this voltage drop controls both the critical current level and the critical voltage of instability.

In experimental conditions, the beginning of the instability practically coincides with the opening of the emitter junction. Therefore, for estimation of the critical regime, the following equation may be used [34]:

$$I_{CR} (R_B + r_B) = 0.7, \tag{1.14}$$

where R_B is the external resistance of the base circuit. With R_B increase, the critical current decreases [34], while the breakdown voltage value U_{CEBR} is practically unchanged. This results in the reduction of the probability of alternative thermal instability mechanisms.

Condition (1.14) covers the case of the base grounded with the external resistor of value R_B as well. In device terminology, this case corresponds to the so-called BVCER breakdown voltage.

2.6.2 The Floating Emitter Case I_E=0

In the case of floating emitter circuit, the injection of electrons becomes impossible. However, from experiment [34], it follows that in a number of power transistors with open emitters, the instability is observed in the $I_E = 0$ regime as well. However, in this case, the critical current level is approximately 10 times higher. The reason for this effect is the increase of the emitter potential with an increase of the residual current in the base up to some critical value. This critical value corresponds to a level of surface breakdown voltage in the region between the emitter and the grounded base.

Due to this specific breakdown scenario, the electrical contact between the emitter and the base is restored and provides a possibility of injection from the emitter junction. In this case, the critical voltage remains the same, but the critical current in reality corresponds to an increase of the emitter potential up to the value U_{EBBR} (which, for the case discussed in [34], was approximately 6 V). Therefore, the critical current for positive feedback and corresponding instability is $I_{CR} = (0.6 + U_{EBBR})/(R_B + r_B)$. Apparently, this condition can be satisfied by the implementation of proper device architecture.

2.6.3 Avalanche–Injection in a Common Emitter Circuit: The Case of $I_B < 0$ Regime

From numerous experimental studies of isothermal instability, it follows that formation of negative differential conductivity in the pulsed regime results in current filamentation in power transistors and ESD devices. The pulse duration t_P is adequate for the filamentation process (for example, $t_P \geq 1$ μs). Therefore, for estimation of the critical conditions, it is sufficient to determine the regimes with $dI/dU \leq 0$.

In the case of active device structures, these conditions can be realized at overloads in the collector circuit in the pinch-off condition. A similar effect can be observed in a large signal operation at load mismatching. However, the most important instances for this book are ESD events.

The application reliability aspect at I_B <0 regime is critical for power transistors. In this regime, the critical drain current level for instability sharply decreases down to less than 1 mA.

The main feature of the negative base current operation is described below. Hole current flow from the base results in the instability of emitter current: the higher the j_B, the higher the value of base potential near the emitter junction. Therefore, the injection level of electrons from the emitter is respectively higher as well. Flowing along the emitter junction (Fig. 2.11) in the base contact direction, the $I_B(y)$ increases emitter current density $j_E(y)$ in the center of emitter strip and dumps it in the edge of the strip. This effect mirrors the Fletcher effect. At fixed emitter current I_E (in common base configuration), the negative base currents might result

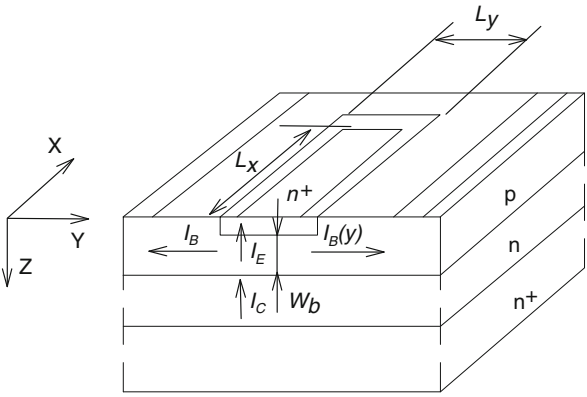


Fig. 2.11 Simplified structure of the power bipolar junction transistor with implanted emitter in case of negative base current operation in common emitter circuit

in filamentation of the emitter current in the "*OX*" direction to the base current flow I_B in the "*OY*" direction.

Under negative base current conditions in the common emitter circuit, the following expression can be used for the base current I_B:

$$I_B = -\frac{I_{C0}}{\alpha} + I_C \frac{(1-\alpha M)}{\alpha M}. \tag{1.15}$$

For sufficiently high potential on the emitter junction, the current is determined by avalanche multiplication in the collector region: $M >> 1$ and $I_B \approx -I_C$. This is similar to the diode circuit for the case $I_E = 0$ discussed above.

For the pinch-off regime, the injection efficiency of the emitter junction is very low, $\alpha << 1$. Therefore, for essential current generation, the conditions $M >> 1$ and $U_{CE} > U_\alpha$ are necessary. The appearance of holes in the base enhances injection and results in emitter current increase. This results in immediate increase of gain α to its maximum value. In this case, high multiplication coefficient values M are no longer necessary to support the already increased current level. At the given current source in the collector circuit, the multiplication coefficient M is decreasing due to U_{CE} reduction. This corresponds to an S-shaped region formation.

The negative differential resistance region was well studied to address reliability problems in NPN BJT. It can be realized at rather small current levels under $d\alpha / dI_E > 0$. At the same time, in structures with high values of the saturation current I_{SE} and I_{SC}, the S-shape may not form at all.

Obviously, from the reliability point of view, this S-shape is not "dangerous," because current redistribution is realized only at rather small current levels and automatically terminated by saturation of the gain α. While filamentation at small current levels does not result in failure, it should be considered that any NDC region on the *I–V* characteristics might be accompanied by spontaneous switching of the

transistor into a new state, according to the load characteristic. This may be critical for operation, as in a typical example of the latch-up scenario.

The I_c increase at $U_{CE} > U_\alpha$ is determined by ratio (1.15). According to the avalanche–injection concept, the critical current level occurs when j_c increases up to the j_0 value (at least locally). At negative base currents ($\alpha M \geq 1$), emitter current redistribution is possible. Therefore, the critical level of current density may be realized at lower I_c levels, in comparison with $I_B = 0$. The typical instability boundary for transistors with this effect is presented in Fig. 2.12a.

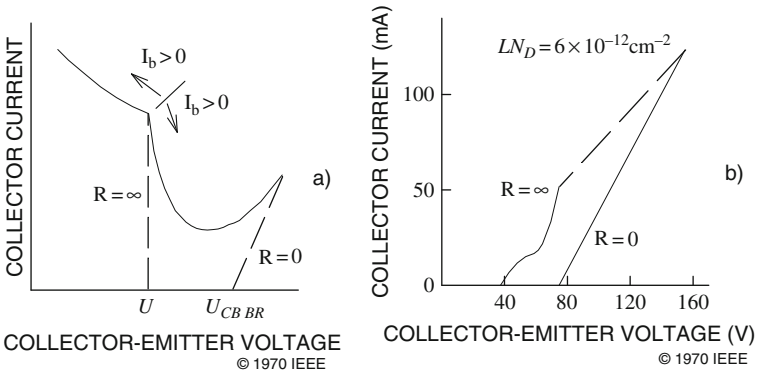


Fig. 2.12 Stability boundary in the negative base current regime ($I_B < 0$) for transistors with collector region parameters $N_D L << 6 \times 10^{11}$ cm^{-2} (**a**) and $N_D L >> 6 \times 10^{11}$ cm^{-2} (**b**) [35]

To demonstrate transistor behavior at negative base current $I_B < 0$, it is useful to analyze the common emitter circuit with resistor R_{BE} between the base and the emitter at a given collector current I_C. This circuit can be used both to measure the pulsed stability boundary U_{CECR} (I_C) at different values of R (from $R_{BE} = 0$ to ∞) and to simulate various base current levels.

A high, but finite, resistance results in a decrease of the gain coefficient, in comparison with the open base circuit case $R_{BE} = \infty$ ($I_B = 0$). Some portion of holes flows out of the base. This corresponds to the condition $\alpha M \geq 1$ and $U_{CE} \geq U_\alpha$ (1.8).

A small negative base current $I_B < 0$ results in the appearance of a double effect. On the one hand, negative base current $I_B < 0$ reduces the emitter injection in comparison with $I_B = 0$ and, on the other hand, it enhances the local current density in the central part of emitter. In the region $U_{CE} > U_\alpha$ (Fig. 2.12a), the redistribution j_C is dominant: while the peak current density j_{MAX} is increasing, the total critical current is decreasing. With R_{BE} decrease, the total injection level is reduced. For the conditions of emitter junction opening, the current increase of I_C ($I_C \approx -I_B$) is necessary:

$$I_{CCR}\,(R_{BE} + r_B) \approx 0.7. \tag{1.16}$$

For a transistor with $N_D L \gg 6 \times 10^{11}$ cm^{-2}, the falling region of I_C at $U_{CE} > U_\alpha$ may not even be presented [35]. Indeed, in such structures, the main part of the collector current I_C is provided by avalanche generation. In this condition, the injection efficiency of the emitter is rather low, so the emitter impact on field redistribution in the n-collector region is negligible. Apparently, in this case, the negative base current flow does not cause any significant redistribution of emitter current. Therefore, the stability boundary has a shape similar to the one presented in Fig. 2.12b. Practically, at $R_{BE} = 0$, the *I–V* characteristic is equivalent to one of the base–collector diode structure.

Evaluation of the critical regime for the avalanche–injection isothermal instability at $I_B < 0$ condition is not as simple as the positive base current case $I_B > 0$. However, analytical estimation for this effect is in rather good agreement with experimental data.

2.6.3.1 Numerical Analysis for $I_B < 0$ Case

In the case of high-gain standard bipolar devices (Chapter 3), the base is much thinner and the collector has a complex profile. Nevertheless, the structure (Fig. 2.13) represents rather practical case of a parasitic device that could both be used for ESD protection and be present in an NMOS device with a submicron gate length. In this chapter, the operation regimes of this important structure are further analyzed at a comparative level of simplification. Integrated NPN BJT devices in analog BiCMOS process technologies with complex profiles and regions are analyzed in Chapter 3, for the cases of both vertical and lateral current transport. This section focuses on essentially the quasi-1D model of the NPN BJT device.

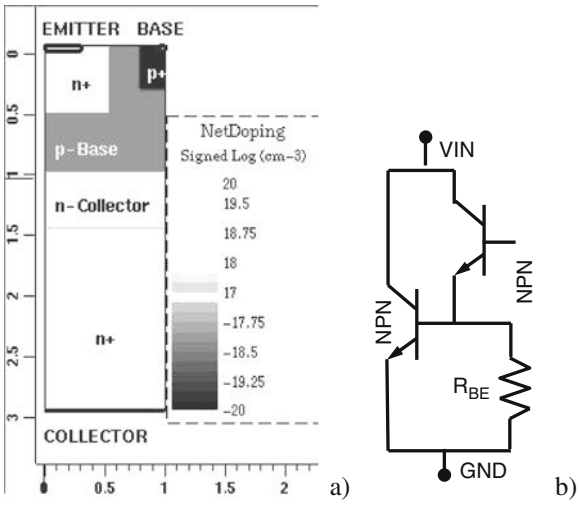


Fig. 2.13 Model of the NPN BJT structure (**a**) and circuit equivalent (**b**) used for numerical analysis and the common emitter device operation

A model of the NPN BJT structure is presented in Fig. 2.13. The structure has three electrodes: base, emitter, and collector. The heavily doped n^+-emitter and n^+-collector regions with $N_D=10^{20}$ cm^{-3} and the p^+-region for base contact with $N_A=10^{19}$ cm^{-3} form contact regions of the structure.

In spite of significant simplification, the internal p-base and n-collector regions are important to preserve in the device structure. The base–collector junction is formed by the n-region with $N_{Dn}=10^{16}$ cm^{-3} and the p-region with $N_{Ap}=10^{17}$ cm^{-3}, respectively.

Conductivity modulation in the BJT structure in avalanche–injection conditions is a strong function of the collector and base profiles. However, the major physical process in the device can be properly illustrated.

The collector–emitter *I–V* characteristics for different base resistances of the NPN BJT structure in common emitter mode are presented in Fig. 2.14 for two base–collector doping compositions:

$$N_{A_p} = 10^{17}\,\text{cm}^{-3}/N_{D_n} = 10^{16}\,\text{cm}^{-3} \text{ and } N_{A_p} = 10^{18}\,\text{cm}^{-3}/N_{D_n} = 10^{17}\,\text{cm}^{-3}.$$

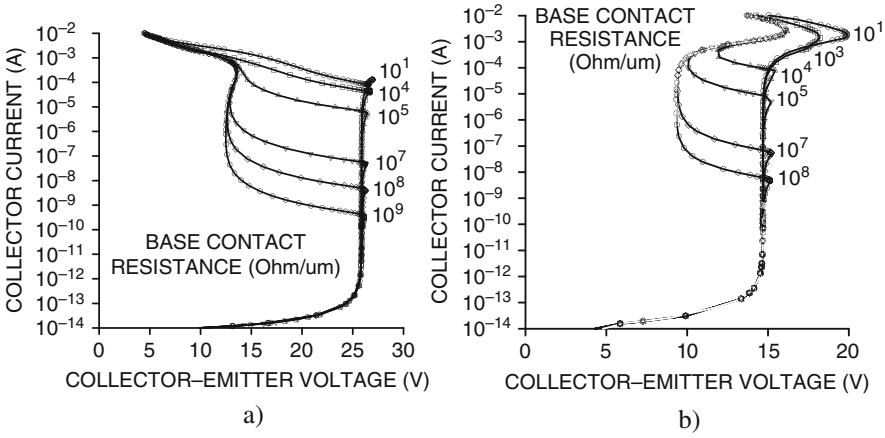


Fig. 2.14 Simulated collector–emitter *I–V* characteristics for two base–collector doping compositions: $N_{Ap} = 10^{17}$ cm$^{-3}/N_{Dn} = 10^{16}$ cm^{-3} (**a**) and $N_{Ap} = 10^{18}$ cm$^{-3}/N_{Dn} = 10^{17}$ cm^{-3} (**b**)

As can be seen from the simulation data, there are two S-shaped regions (Fig. 2.14) that can be identified in the output collector–emitter *I–V* characteristics.

After avalanche breakdown and small current saturation, the structure exhibits negative differential resistance with two negative differential conductivity regions.

At the first critical triggering point of instability V_{TA} (Fig. 2.14a), the device reaches the condition with a minimum holding voltage V_{HA}, followed by a new current saturation region that brings the voltage drop to a new triggering point of instability V_{TB}.

For the first current instability region, the dependence upon base contact resistance remains in good agreement with the analytical expression (1.16) for silicon material at room temperature conditions $I_{CR}\,(R_B + r_B) = 0.8 = \text{const}$.

However, this correlation is mainly true only at rather high base resistance and respectively relatively low critical current below 0.1 mA/μm (Fig. 2.14b).

In case of strong injection in the base–collector region, additional effects become involved. These effects are current saturation and conductivity modulation of the collector region itself.

To explain these phenomena, the *I–V* characteristics of the NPN structure with $N_{Ap}=10^{18}$ cm^{-3}/$N_{Dn}=10^{17}$ cm^{-3} and base contact resistance $R_{BE}=10^4$ Ω/μm (Fig. 2.15a) are compared to the dependence of base contact bias and negative base current (Fig. 2.15b).

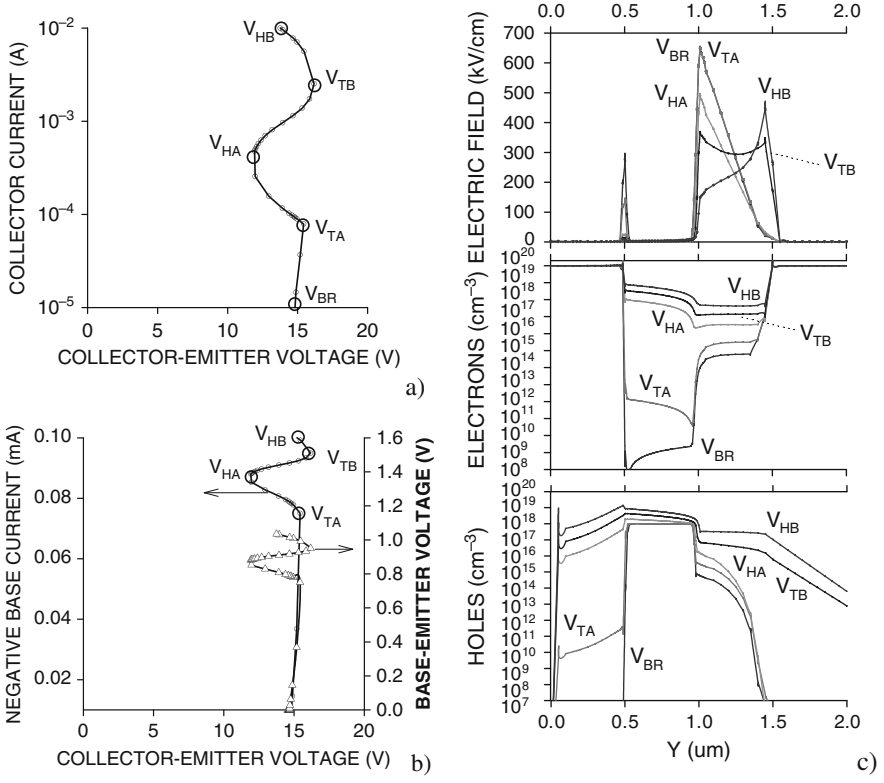


Fig. 2.15 Conductivity modulation of the base and collector regions in the NPN BJT structure model. Dependence of the collector current (**a**), negative base current (**b**), and distribution of the electric field magnitude and electron and hole density (**c**) in the selected conductivity modulation regimes V_{BR}, V_{TA}, V_{HA}, V_{TB}, V_{HB}

In comparison of these data in the range before the instability point V_{TA}, the negative base current is practically equal to the collector current. In the first current instability point V_{TA}, the base contact voltage is reaching ~0.8 V, thus eliminating the potential barrier from the n^+-emitter region injection into the p-base region, for the case of room temperature silicon material.

In the V_{TA} regime, the distribution of the electric field magnitude corresponds to the peak at the collector–base junction with the space charge (Fig. 2.15c, plot for the electric field distribution).

With the following current increase, the density of the injected carriers in the n-collector region exceeds the donor concentration in this region (Fig. 2.15c, distribution for electrons) and results in the current saturation effect in this region. Due to a much higher current in the base–collector region, current conduction is supported at a much lower multiplication coefficient under lower electric fields (Fig. 2.15c, distribution of electric field in regime V_{TB}).

At further current increase, modulation of the n-collector region conductivity is observed. The density of injected carriers in the n-collector exceeds the N_{Dn} level. This effect results in electric field redistribution with the maximum of the dependence formed at the n–n^+ collector interface. With the higher current density, the quasi-neutral region is expanded toward the collector n^+–n junction, thus providing a lower voltage drop on the structure, which is associated with the second-stage negative differential resistance.

In these conditions of strong injection, the injected carriers density in the structure is higher than $N_{Ap}=10^{18}$ $cm^{-3}/N_{Dn}=10^{17}$ cm^{-3} and the device operation is similar to the high-current operation of the n^+–p–n^+ structure with a floating base, where conditions on the base provide practically no effect (Fig. 2.15b).

2.6.4 Avalanche–Injection in the Common Emitter Circuit with Positive Base Current $I_B > 0$

The common emitter NPN-based ESD clamp circuit is perhaps the best known ESD protection configuration. In the case of ESD clamp, the positive base current can be provided by an additional reference clamp component. This component can be represented by an avalanche diode with an appropriate breakdown voltage, an open base NPN BJT with an appropriate BVCEO voltage, or another device with breakdown type *I–V* dependence. An alternative solution can be realized with a more complex control circuit that will provide a positive base current according to a more sophisticated algorithm that accounts for conditions on other electrodes and different time delays (Chapter 3). In appropriate design of the reference circuit, the snapback clamp can eliminate the transient time-dependent turn-on effect (d*V*/d*t* effect).

This mode of operation is discussed using both the analytical and numerical simulation approaches for the structure (Fig. 2.13).

2.6.4.1 Analytical Description of the $I_B > 0$ Case

For simplification, the base current in the common emitter biasing circuit is assumed constant and positive: $I_B = const,\ I_B > 0$. Neglecting the saturation current contribution, in comparison with the high base current, the collector current is equal to

$$I_C = \frac{\alpha M I_B}{1 - \alpha M}. \tag{1.17}$$

Since the multiplication coefficient is the function of the collector–base voltage and the collector current $M = M(U_{CE}, I_C)$, (1.17) determines the device *I–V* characteristic. By differentiating the right part of (1.17) on U_{CE} at $\alpha = \text{const}$, $T = \text{const}$, and $\partial M \big/ \partial U_{CE} > 0$, the simple criterion of instability is provided by

$$\frac{I_C}{M\,(1 - \alpha M)} \frac{\partial M}{\partial I_C} > 1. \tag{1.18}$$

Unfortunately both criteria (1.18) are obtained involving both the current gain $\alpha(I_C)$ and the multiplication coefficient $M(T)$ dependencies. Thus, the criterion is hard to use in practical cases since it requires the exact dependencies $\alpha(I_C)$ and $M(I_C, U_{CE})$.

For some particular cases, it is possible to complete the calculation of the instability boundary. Meanwhile, the result from this calculation is no more informative than the qualitative estimation presented below.

At low base currents when $j_C < j_0 = qN_D v$, the transistor behavior is simply similar to the case of $I_B = 0$ presented above. In the vicinity of U_α, the collector current begins increasing according to (1.17) up to the critical value I_0, when the negative differential conductivity is formed, followed by the current instability.

At a corresponding base current level, when the collector current $j_C > j_0$, the maximum electric field is formed on the subcollector n–n$^+$ junction. Resulting conditions are formed for avalanche–injection instability immediately with avalanche generation, i.e., at the electric field $E_{MAX} = E_{BR}$. This effect is observed at $U_{CEBR} < U_\alpha$, where U_α is the breakdown voltage in the CE bias configuration derived from the condition $\alpha_1 M(U_\alpha) = 1$.

It is quite clear that with the increase of I_C, a consequent value of U_{CEBR} will be decreased. For $I_C > 2I_0$, the thickness of the space charge region W will still be lower than the thickness of the n-epilayer L. With the assumption of triangular electric field distribution, the integral value of the collector–emitter voltage in the critical point is $U_{CECR} \approx E_{BR} W_{CR} \big/ 2$, where $W_{CR} \approx (E_{BR}\varepsilon) \big/ \left| j \big/ v - qN_D \right|$. Therefore, for adequately high current $I_C \gg I_0$, the isothermal stability boundary must be close to the hyperbolic function:

$$U_{CR} I_{CR} \approx \frac{\varepsilon E_{BR}^2 v}{2}. \tag{1.19}$$

2.6.4.2 Numerical Analysis of the $I_B > 0$ Case

In order to compare with the analytical approach, the same NPN BJT device structures with $N_{Ap} = 10^{17}$ cm^{-3}/N_{Dn}=10^{16} cm^{-3} and N_{Ap}=10^{18} cm^{-3}/N_{Dn}=10^{17} cm^{-3} (Fig. 2.13a) are solved with current boundary conditions for the base contact. The same silicon parameters, room temperature, and quasi-static solution have been calculated.

At different positive base currents, it can be observed that the instability point depends on the device design that provides different current gain. The structure with parameters N_{Ap}=10^{17} cm^{-3}/N_{Dn}=10^{16} cm^{-3} in Fig. 2.16a provides current gain of over 1, with dependence close to hyperbolic (1.19), while the device with parameters N_{Ap}=10^{18} cm^{-3}/N_{Dn}=10^{17} cm^{-3} has gain lower than 1 and provides a different type of collector–emitter characteristic (Fig. 2.16b).

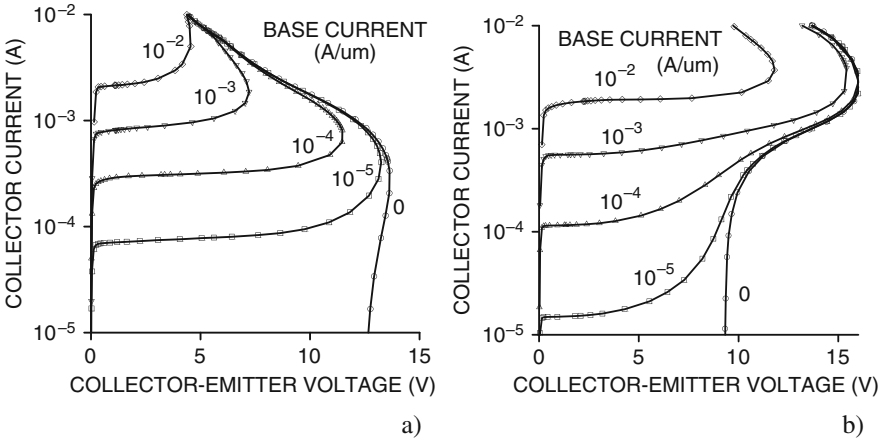


Fig. 2.16 Simulated collector–emitter *I–V* characteristics for two base–collector doping compositions: N_{Ap}=10^{17} cm^{-3}/N_{Dn}=10^{16}cm^{-3} and N_{Ap}=10^{18} cm^{-3}/N_{Dn}=10^{17} cm^{-3} at positive base current

This example demonstrates that numerical simulation analysis provides a more accurate and simple insight into device characteristics than a simple analytical estimation.

In the device with low current gain, completion of the condition $\alpha M > 1$ requires much higher multiplication coefficients, and thus a much weaker dependence upon base current is observed.

While high-current gain is one of the major figures of merit for NPN devices supported in integrated process technology, low current gain structures are not rare cases. Most practical examples of such occurrences are provided by parasitic structures formed in NMOS devices, as well as by parasitic structures that might be formed between different devices in the drawn layout, for example, between the

n^+-source and n^+-region of the antenna diode in Pwell of the NMOS device (see Chapter 6).

Using numerical simulation, the physics of conductivity modulation in a structure with positive base current can be understood by comparing the electric field and carrier densities in the conditions V_{TA} and V_{HA} (Fig. 2.17a).

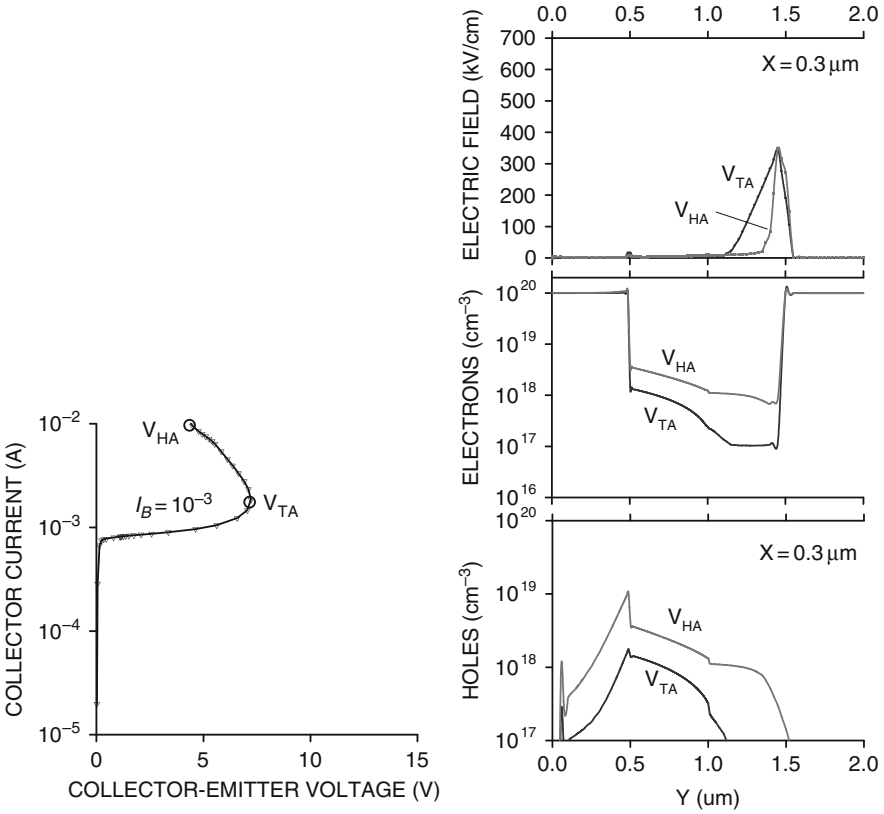


Fig. 2.17 Collector–emitter *I–V* characteristics of the NPN structure (Fig. 2.13a) with the base–collector doping level $N_{Ap}=10^{17}$ cm^{-3}/$N_{Dn}=10^{16}$ cm^{-3} under positive base current $I_B=10^{-3}$ A and distribution of the electric field magnitude, electron and hole density for cutline $X = 0.3$ μm in the triggering and holding regimes V_{TA} and V_{HA}

Similar to n^+–p–n^+ structures, discussed above, negative differential resistance is formed due to the reduction of the electric field in the n-collector region between $Y = 1$ and 1.5 μm (Figs. 2.13a and 2.17b). Thus, namely conductivity modulation of the n-collector region is responsible for the formation of S-shaped characteristics in this structure.

2.6.5 Avalanche–Injection in the Common Base Circuit

The common base circuit operation has relatively rare ESD applications. However, for completeness of the presented material, this case is discussed in this section based upon the results of early experimental studies.

The conditions of isothermal instability in the common base bias circuit (CB) and common emitter bias circuit (CE) are slightly different from each other at $U_{CB} < U_\alpha$. For a current density above the critical level $qN_D v$, the instability begins immediately with avalanche–injection at U_{CB} of avalanche ionization:

$$U_{CB} \approx \varepsilon E_{BR}^2 \Big/ 2\left(\frac{j_C}{v} - qN_D\right).$$

At $j_C > 2j_0$, the device turns on into the negative base current regime. For CE circuits, this effect is accompanied by the loss in collector control by the base current. For CB circuits, the voltage range $U_\alpha < U_{CB} < U_{CBCR}$ is still operational.

The negative base currents may result in a redistribution of the emitter current. In CE circuits, this redistribution involves a part of injected electrons from the emitter followed by a redistribution of the total collector current that has an avalanche component. In CB circuits, at $U_{CB} < U_{CBCR}$ the collector current is practically equal to the emitter current ($j_C = \alpha M j_E$). Therefore, the degree of current redistribution might be significantly higher. At $I_E = const$, it can result in a scenario where the isothermal filamentation of the emitter current is not translated to the external circuit as negative differential resistance, but essentially takes place under positive differential conductivity $I\,(U_{CB})$.

At $U_{CB} > U_\alpha$, the hole space charge is increasing the forward bias level of the emitter junction. The non-uniformity of this process results in redistribution of emitter current as follows: the higher the hole density, the higher both emitter forward bias and the emitter current density. When flowing through the avalanche multiplication region, the density of the electron current results in an increase of the avalanche hole density. If in these conditions the current density j_{CMAX} in the middle of filament becomes equal to the critical value j_0, then the triggering-on is observed according to the classical avalanche–injection mechanism.

From a theoretical point of view, the isothermal instability in CB might have a two-stage evolution. The evolution of avalanche–injection in the collector n-layer is not a necessary condition for transistor failure. Emitter current filamentation at $\alpha M > 1$ cannot cause the catastrophic consequences by itself, because the total dissipated power $P_C = U_{CB} I_C$ is practically unchanged, namely, the current redistribution results in a sharp increase in maximum specific power $j_{CMAX} U_{CB}$.

A distinct feature of the isothermal instability in CB at real operation is the low level of excessive currents before voltage redistribution (at least one order of magnitude lower than in the CE bias circuit). Critical excessive current ΔI_{CR} means that there is a difference between the values of I_{CR} and I_C before the NDC formation (Fig. 2.18a). In CE, the value of ΔI_{CR} may be achieved at ~0.1 mA, but in CB ΔI_{CR} in the $U_\alpha < U_{CB} < U_{CBCR}$ range it is higher by an order of magnitude.

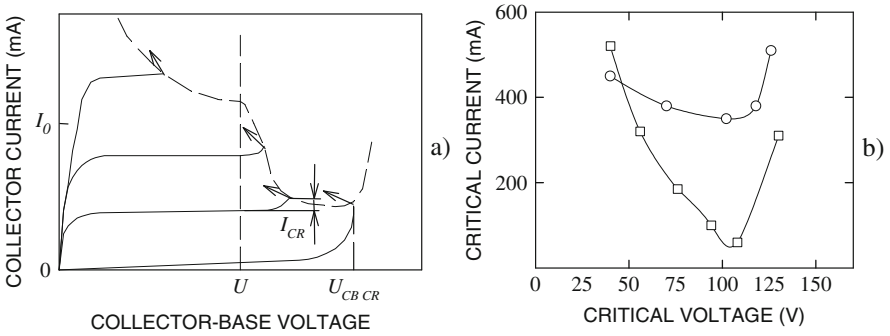


Fig. 2.18 Isothermal instability in the common base circuit: *I–V* characteristic (**a**) and experimental stability boundary for two samples (**b**)

The experimental shape of the stability boundary (Fig. 2.18b) corresponds to a concept of primary emitter current filamentation. The decrease of I_{CR} at $U_{CB} > U_{\alpha}$ is apparently the same as the decrease of a similar value in CE circuits. It should be considered that the value varies for various devices (of a given transistor type) and is the result of device parameter scattering and presence of local structural defects (Fig. 2.18b). This problem will be discussed in the following section for filamentation criterion.

2.7 Avalanche–Injection in PNP Structures

ESD clamps based upon the p–n–p structure are key components for ESD protection solutions for fast transient pins, hot-plug-in, and system-level applications.

In principle, the processes in p–n–p BJT structures can be described using the same analytical approach that is presented above for the other NPN structures. However, the major difference in the case of silicon-based process technologies is implied by the difference in carrier mobilities and multiplication coefficients.

This difference in parameters for electrons and holes in the case of real device parameters provides a much higher holding voltage and much smaller negative differential resistance. In addition, the positive feedback due to avalanche–injection processes has a much higher compensation by the negative feedback of lower conductivity in the case of holes as majority carriers.

At the same time, a mild negative differential resistance is a rather useful device feature that provides an ideal vertical *I–V* dependence of the clamping voltage upon current by compensation of the higher resistance of the p-regions (Fig. 2.19).

Practical realization of both effects requires substantial accuracy. Therefore, numerical simulation of such p–n–p structures presents a valuable advantage. To demonstrate the difference between silicon n–p–n and p–n–p structures, a simulation was completed for a similar PNP BJT structure and is presented in Fig. 2.20.

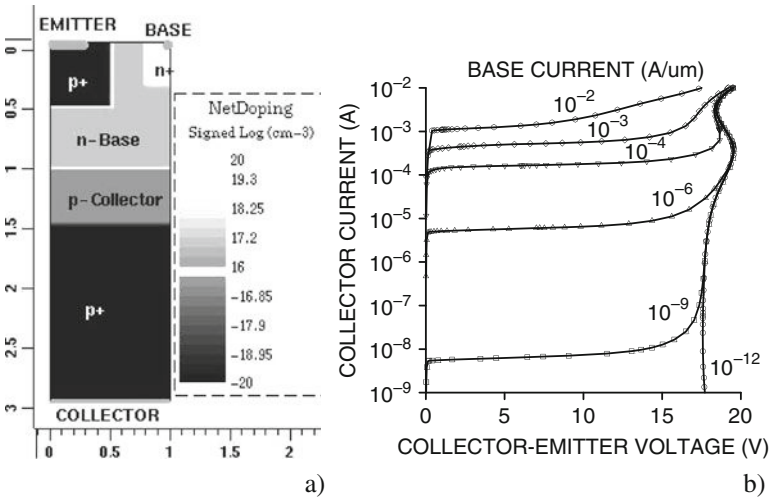


Fig. 2.19 Simulated collector–emitter *I–V* characteristics for the PNP structure with reversed doping type in comparison with the NPN structure (Fig. 2.13a) for different base current levels in the common emitter circuit

The 0.5 μm long p-collector and 0.5 μm long n-base with a uniform doping profile were defined similar to the corresponding levels in the NPN device (Fig. 2.16a) $N_{Dn}=10^{17}$ cm^{-3} and $N_{An}=10^{16}$ cm^{-3}, respectively. The base current direction and collector voltage have been, respectively, reversed.

According to the cutline plots, namely modulation of the n-base provides for the S-shaped dependence. The distribution of the electric field and carrier density across the cutlines at $X = 0.3$ μm is presented in Fig. 2.20.

Thus, this simple numerical experiment provides several practical guidelines for PNP-based ESD clamp design (Chapter 4).

2.8 Double Injection in Si p–n–p–n Structures

2.8.1 Equivalent Circuit

This section deals with structures that are key components of power electronics and parts of many power management devices: insulated gate bipolar transistors (IGBT), silicon-controlled rectifiers (SCR), DIACs, TRIACs, and many others [14]. In addition, these elementary structures represent the parasitic devices in many different latch-up scenarios.

The basic theory of thyristor can be found in [14]. This section is focused mainly on p–n–p–n structures specifically related to ESD application and latch-up.

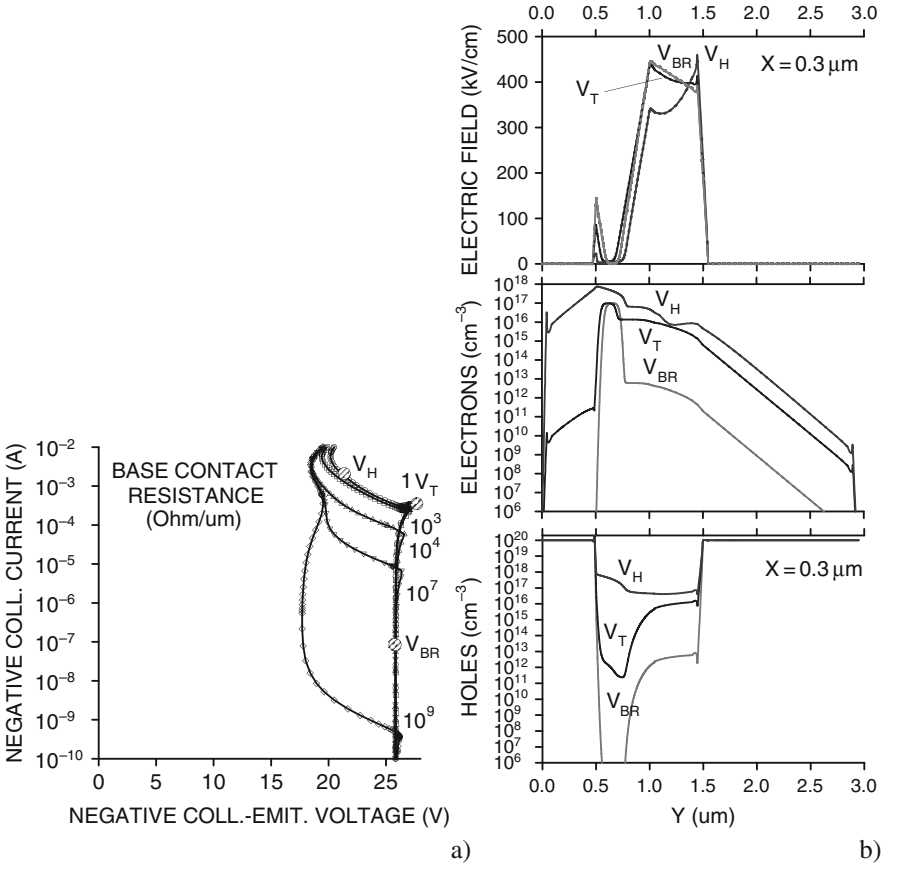


Fig. 2.20 Calculated collector–emitter *I–V* characteristic for the PNP structure with different lumped base resistances (**a**) and comparison of the electric field and carrier density distribution for the cutline at $X = 0.3$ μm in representative regimes (**b**)

As well, the region parameters are chosen typical of 0.5 μm process technologies. Conductivity modulation effects in such devices are demonstrated mainly by numerical simulation.

The final stage of the physical mechanism of positive feedback realized in p–n–p–n structures is called double-injection conductivity modulation. A distinct feature of such a modulation mechanism is that it can be realized at rather low voltages of ∼1.5 V. In these low electric field conditions, avalanche breakdown in the structure is suppressed and not required to support conduction of high current, unlike in the case of conductivity modulation mechanisms discussed above.

Due to low holding voltage in high-current regime, the dissipated pulsed power is low as well. This provides a significant advantage for ESD design that can be taken from both low dissipated power in the active region and the low clamping voltage itself. In a properly designed device, the double-injection conductivity modulation

provides internal current density over a 10 mA per micron width. This provides a very important device-level solution for high-speed applications, due to such a low parasitic capacitance required for a standard package-level spec.

The low (below 2 V) clamping voltage has a critical advantage over protection by a leaky diode stack for nanoscaled CMOS processes.

However, the low clamping voltage has a downside as well. In case of high-voltage and transient pins, where transient latch-up conditions can be created, it is hard to realize SCR devices with holding voltage above the power supply level. Several solutions for this will be discussed in the next chapter.

In some cases, the p–n–p–n structure can be analyzed as made of four layers with a corresponding sequence of the p^+-emitter, n-base, p-base, and n^+-emitter.

Most of the physical effects in the device can be understood using a so-called Ebers circuit [14] for device representation (Fig. 2.21). In this structure, the n–p–n and p–n–p devices are sharing corresponding collector and base regions. In submicron dimensions, connection of the bases is necessary to suppress gain of the parasitic n–p–n and p–n–p in the structure (Fig. 2.21a).

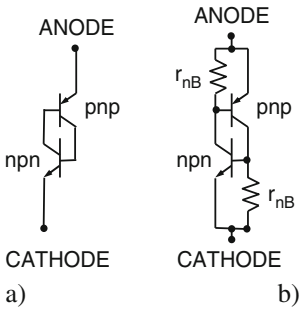


Fig. 2.21 Ebers circuit used to represent the p–n–p–n structure with open external bases (**a**) and shorted bases (**b**)

Therefore, the correct way to study conductivity modulation in the p–n–p–n structure requires the use of a version where the n-base and p-base are connected. The most typical circuit configuration for this structure in submicron ESD designs is presented in Fig. 2.21b. This parasitic structure can be formed by CMOS inverter circuit (PMOS and NMOS) in the following sequence: p^+-source and n-well of PMOS and p-well and n^+-source of NMOS.

A two-terminal diode structure is realized in the case of floating base regions (Fig. 2.21a). In general, the blocking junctions can be connected to a reference voltage and current subcircuit in order to realize a desired level of the triggering current and voltage (Chapter 3).

One of the particular cases realized in submicron process technologies for ESD protection is device implementation with a p–n–p–n structure where the base regions are connected to the corresponding anode and cathode regions (Fig. 2.21b). The most useful operation regime for ESD protection is one where the p^+-contact region

is positively biased. In this case, the blocking junction is created by the high-voltage-tolerant n-base to p-base junction.

The phenomenological principle of device operation can be understood using the theory of bipolar devices and the equivalent circuits (Fig. 2.21).

Criteria for positive feedback that results in conductivity modulation in p–n–p–n structure can be obtained by

$$\alpha_{\mathrm{NPN}} M_N + \alpha_{\mathrm{PNP}} M_{\mathrm{p}} > 1, \tag{1.20}$$

where α_{NPN}, α_{PNP}, M_{N}, $M_{\mathrm{P,}}$ are current gain and multiplication coefficients in the internal n–p–n and p–n–p structures.

In the case of floating base regions (blocking junction contacts), the p–n–p–n structure operation depends upon the formed SCN or SCL structures.

The collector current of each structure provides for the base current of the other structure. Thus, even if one of the devices is achieving the condition of αM >1, instability is unavoidable.

In submicron process, the base regions are usually rather small and comparable with diffusion length. In this case, a more practical ESD device is realized using a connection of the blocking junction regions (Fig. 2.21b).

2.8.2 Simulation of Conductivity Modulation in p–n–p–n Structures

A corresponding numerical solution can be obtained using a 2D structure with the base layers made of a sequence of the anode p^+-region, n-base, p-base, and cathode n^+-contact region.

The conductivity modulation processes in p–n–p–n structures with floating and connected base regions are demonstrated by means of numerical simulation in Figs. 2.22 and 2.23, respectively.

2.8.2.1 Floating Base Case

On the phenomenological level, the principle of conductivity modulation in p–n–p–n structure operation is rather easy to understand using simulation results for different doping levels in the base regions.

In the four-layer device with floating blocking junction regions (Fig. 2.22a), when the p^+-anode is positively biased, the current in the device is blocked by the reverse-biased junction that is formed by the n–p bases.

With voltage increase, the critical voltage depends on the space charge region length as compared to the base region length. At a few volts drop on the structure, in the case of low base doping of $<10^{16}$ cm^{-3}, the space charge region covers the whole 1 μm base region. This provides a scenario for current conduction mode under high-current gain conditions with α_{NPN}, α_{PNP}>>1 under small M_{N}, M_{P}∼1. Significant

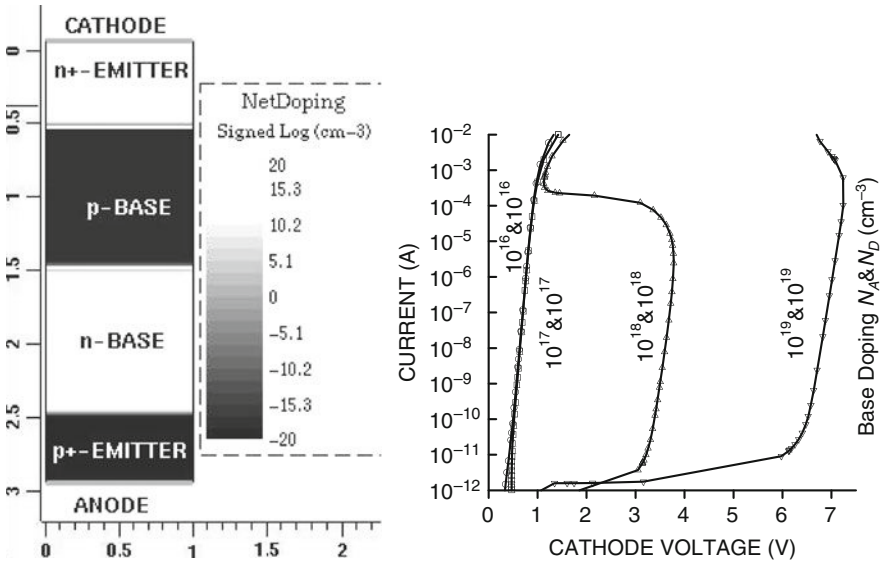


Fig. 2.22 Two-dimensional model for simulation of the conductivity modulation effect in p–n–p–n structures (**a**) and simulated isothermal quasi-static *I–V* characteristics with different doping levels in the base regions (**b**)

amplification of the saturation current at low voltages results in conductivity modulation (1.20) at very small multiplication coefficients. With further current increase, the high-conductivity region is formed (Fig. 2.22 for $N_{Ap}=N_{Dn}=10^{17}$ cm^{-3}).

However, at higher doping levels of the base regions, the critical voltage for conductivity modulation rises. In this case, the space charge region length is less than the base length, and thus the punch-through operation is not realized. This enables an increase of the electric field at the blocking junction up to a level suitable for avalanche breakdown. This provides high multiplication coefficients M_N or $M_P >> 1$ and corresponding multiplication of the saturation current and low current gains with completion of condition (1.20).

At higher current, the saturation of the n and p regions is observed and results in expansion of the space charge region up to the contact n^+ and p^+ regions. After creation of the bias of ~0.7 V, the potential barrier is eliminated and injection from the corresponding p^+–n and n–p^+ forward-biased junctions begins. This forms a positive feedback that creates the double-injection conductivity modulation mode.

In high injection mode, the injected high-density electron–hole plasma overmodulates the conductivity of both base regions. Due to positive feedback, the structure can support high current conduction without high avalanche multiplication $M_N, M_P \sim 1$.

That is why this type of conductivity modulation is called double injection, in opposite to avalanche–injection as in n–p–n structures, where the positive feedback

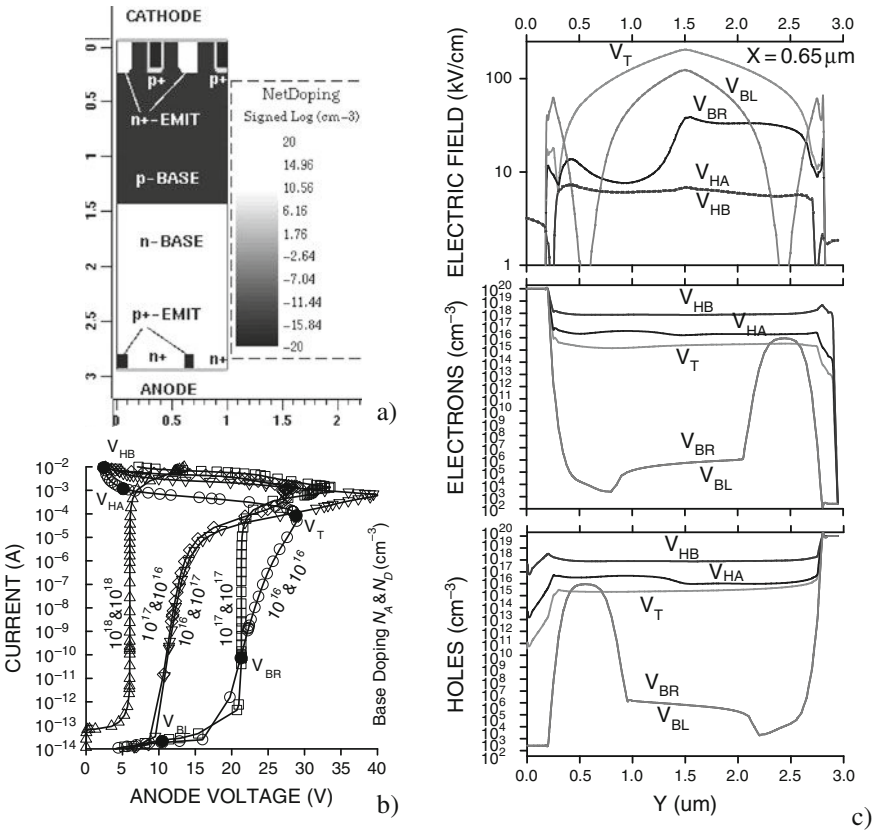


Fig. 2.23 Two-dimensional model for simulation of the conductivity modulation effect in p–n–p–n structures with shorted blocking junctions (**a**); simulated isothermal quasi-static *I*–*V* characteristics at different base doping levels (**b**) and electron and carrier density distributions for different current with different doping levels in the base regions (**c**)

is realized due to low-field injection from the emitter n^+–p junction and avalanche generation at the collector p–n^+ junction. Respectively, this type of conductivity modulation in reverse-biased p–i–n is called double-avalanche–injection due to the positive feedback from two multiplication regions formed in the structure.

2.8.2.2 Connected Base Case

The simulation example for a two-terminal device with connected base regions is presented in Fig. 2.23. To simplify the example, corresponding interdigitated n^+ and p^+ base contact regions are defined in the structure.

The upper and lower contact regions are connected to the corresponding base and emitter regions, thus forming a diode structure.

The final two-terminal device is a good physical representation of a real ESD device implemented in integrated process technologies. The scenario is discussed above, with corresponding correction to the operation of the internal n–p–n and p–n–p structures with negative and positive base currents, respectively.

This mode of operation provides corresponding high blocking voltages at low base doping levels, by reducing the current in the multiplication region due to base connection to the anode and cathode.

Since the bases are externally connected, the saturation current has an escape path. In this case, even at low base doping, the structure is capable of achieving high breakdown voltage. Only when avalanche breakdown followed by the avalanche–injection in either n–p–n or p–n–p provides a sufficient current level, the double-injection conductivity modulation becomes dominant. This conductivity modulation provides positive feedback that terminates avalanche breakdown. In the final state, the device is in on-state and conducts current similar to the forward-biased p–i–n diode, since the injected carrier density significantly exceeds the doping level in the base regions. In Chapter 3, various SCR devices based upon the p–n–p–n structure with double-injection conductivity modulation will be discussed in greater detail.

2.9 Spatial Current Instability Phenomena in Semiconductor Structures with Negative Differential Resistance

The formation of spatially non-uniform states in distributed systems can be observed at certain conditions of different spatial parameters for positive and negative feedback processes. The parameters for these conditions are internal to the system. There are many examples of such systems and phenomena in traffic, biological, chemical, social, and semiconductors [26].

Usually, the current density instability processes in the active device can be described physically using two parameters-activator and inhibitor. Spatially non-uniform states can be formed such that the inhibitor and activator have different characteristic spatial or temporal parameters [26].

In chemical systems, these internal processes can be represented by reaction and diffusion processes [23–27]. In the case of semiconductor systems, current instability with a small spatial parameter acts as an activator, while a diffusion or current spreading in the quasi-neutral areas acts as an inhibitor.

As for distributed systems, the case of ESD devices practically always presents a distributed object where the required total current is achieved by appropriate width scaling.

Practically all elementary semiconductor structures, as described above, provide at certain conditions an example of the system where interplay between positive and negative feedback mechanisms provides for formation of spatially non-uniform current conduction.

The activation process in the structure is the avalanche–injection process itself that provides positive feedback on the local level. Usually, the spatial dimensions of

such a parameter are approximately equal to the space charge region of the structure where the conductivity modulation is realized.

The inhibition effects are provided by the ballasting and contact regions in the structure, due to current spreading to such regions. Thus, the typical spatial dimension of the inhibitor is roughly equal to the saturation region's length.

In the ESD specific case, the primary current filamentation effect, due to conductivity modulation of the discharge regions during the snapback, is controllable and reversible. The secondary filamentation effect is usually responsible for damage of the ESD devices in the pulsed mode.

This phenomenon is similar to the physical limitation of any device [9]. Practically, spatially uniform damage is never observed in distributed semiconductor structures. On the contrary, the most typical scenario is a solitary or multiple-structure melting region (Fig. 2.24). However, often the original damage which occurred in pulsed regime might be significantly altered by the subsequent pulse zaps or functional tests. Often, the following EOS event which occurs under power-on conditions creates a more significant damage due to a domino effect.

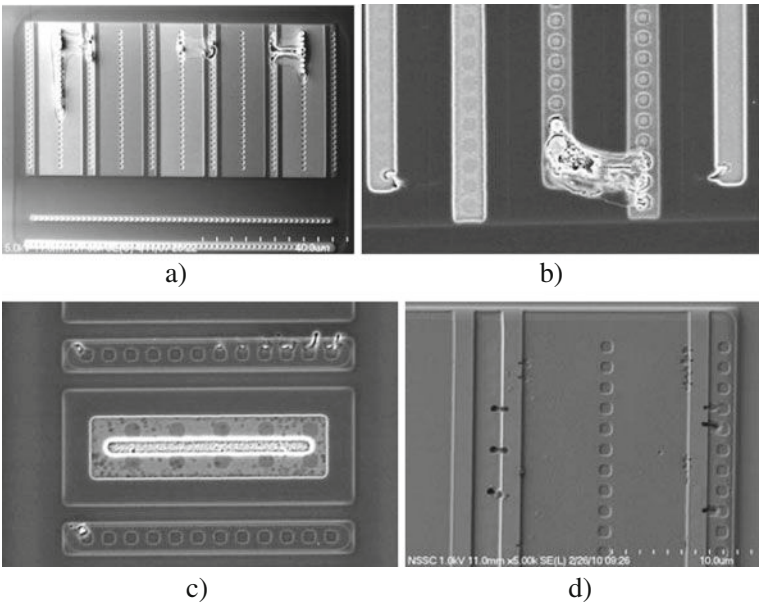


Fig. 2.24 Examples of failure analysis photos for catastrophic burnout in the ESD snapback NMOS device (**a**), ESD diode (**b**) NLDMOS as result of HBM (**c**) and MM (**d**) pulses

2.9.1 Current Filamentation at Avalanche–Injection

The first 2D numerical simulation of the avalanche–injection current filamentation effect has been presented in [15] in order to explain the physical nature of the drain-

source burnout effect in MESFETs [16–19]. The phenomenon has been identified in the parasitic n-i-n bipolar device.

Similarly, the avalanche–injection current filamentation effect has been simulated in [20] for silicon n–p–n structures.

To model the base region connection with 2D numerical simulation capabilities, a multicell mathematical approximation structure has been suggested (Fig. 2.25a). Thus, an essentially 3D problem has been solved within 2D numerical simulation capability by the composition of a 2D triode n^+–p–n^+[p^+] structure. The structure

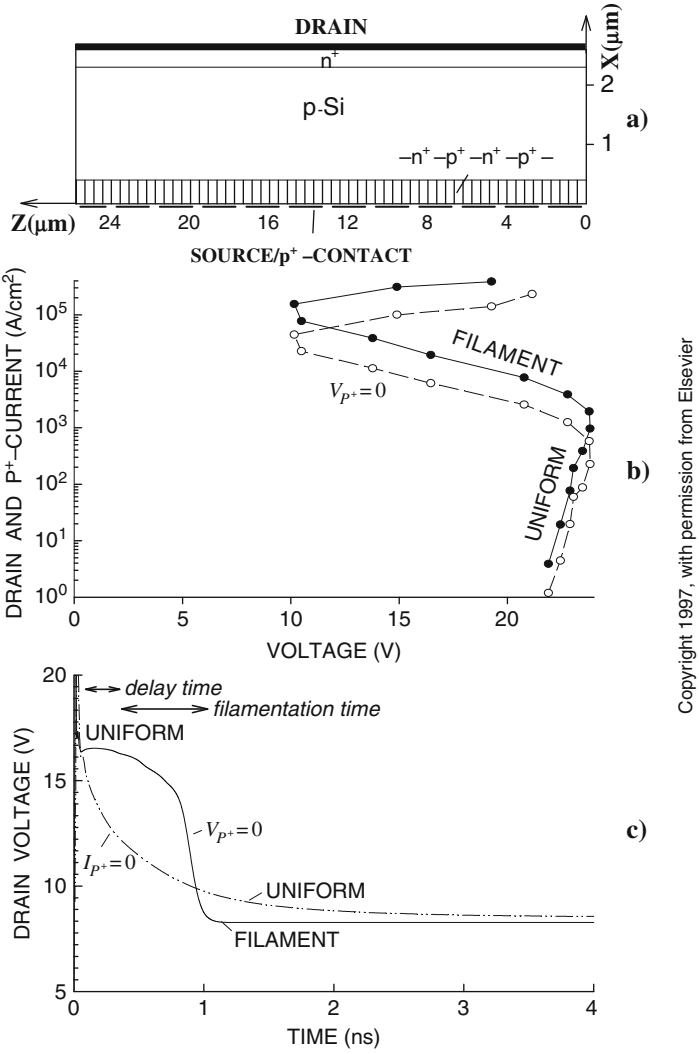


Fig. 2.25 Multicellular n^+–p–n^+[p^+] structure (**a**), calculated I_D–V_{DS} characteristics for the grounded and floating p^+-cell configuration (**b**), and the drain voltage vs. time after a stepwise current increase from $I_D = 0$ up to $I_D = 20$ kA/cm^2 (**c**) [20]

presents itself as a periodic structure with multiple n^+- and p^+-cells of the source and substrate contact regions.

In order to reduce the impact of the spatial structure of the device regions on the spatially non-uniform solution, the dimensions of n^+- and p^+-cell contacts have been chosen smaller than the expected filament dimension. An iterative method can be used in the simulation to demonstrate that at certain dimension of the contact regions, there is no observed influence on the final filament state. The structure can be understood as a distributed cross section of low gain multifinger NPN BJT, where emitter and base regions are reduced below reasonable process limits and butted, while the p-base region is rather big.

A numerical solution has been obtained using transient isothermal simulation for the conditions of applied constant current.

From general understanding of the non-linear theory [23–27], in order to initiate the stratification of the solution into a spatially non-uniform state, small fluctuations of the uniform distribution are needed. While in real devices the fluctuations are always present due to statistical effects and non-uniformities, a physical equivalent for fluctuation is required in the numerical simulation of the ideal structure to enable the possibility of current stratification. According to Vashchenko et al. [18] numerical error of solution provides an adequate source of fluctuations. As a result, this spatial current instability effect can be simulated [18]. This numerical "noise," due to finite accuracy of the numerical solution, provides conditions for filament excitation with no additional fluctuation.

As has been demonstrated [20] (Fig. 2.25b), the transient solution for the distributed structure (Fig. 2.25a) was spatially non-uniform. The initial current density applied at the beginning of the transient solution was chosen at the level corresponding to the negative differential resistance in a minimum contact width device. After a fast transient process, the first unstable uniform state is formed during a very short period of time (<1 ns) (Fig. 2.25c), and then the final quasi-static solution corresponds to the current filament. The distribution of the electric field and carrier density for the current filament state is presented in Fig. 2.26.

Within the filament region, the peak of the electric field peak at the drain (collector) junction (Fig. 2.26a) corresponds to avalanche–injection conditions at high current, with corresponding quasi-neutral electron–hole plasma in the p-region of the structure (Fig. 2.26b, c).

The amplitude of the filament is roughly limited by the current saturation effect in the contact n^+-regions. Therefore, further current increase results in broadening of the filaments up to the structure width.

The important practical conclusion from this numerical experiment is the effect of the contact regions in ESD devices. If the n^+-region is inadequate, excessive injection from the contacts and power dissipation in the metal–semiconductor interface results in very fast local damage of the structure. This effect is significantly amplified due to a much lower melting temperature of the contact eutectic of $\sim$600°C in comparison with melting effects in the semiconductor material of $\sim$1100°C.

The technique of the current limitation in conductivity modulation mode is practically realized in many ESD protection structures, including snapback NMOS devices (Chapter 3).

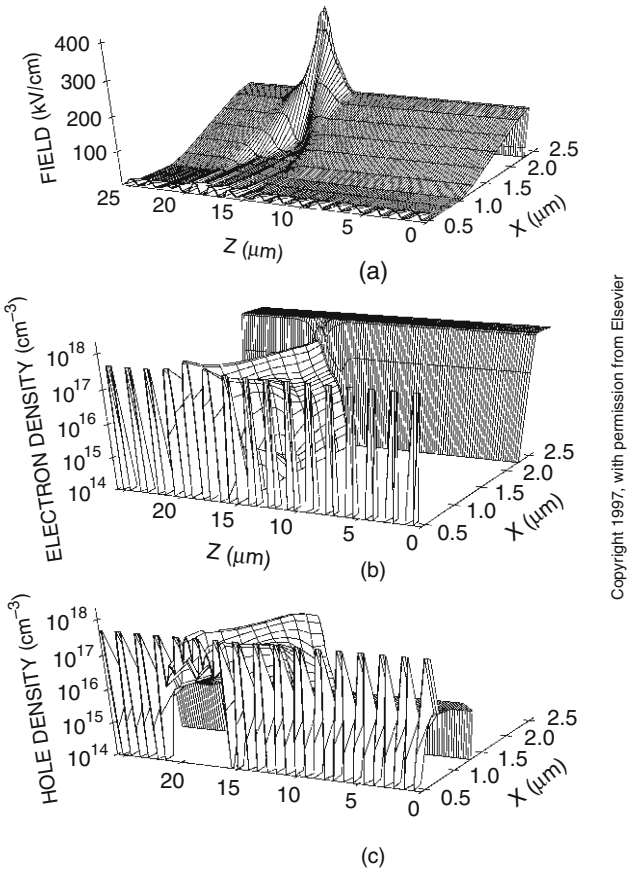


Fig. 2.26 Calculated distribution of the electric field (**a**), electron (**b**), and hole (**c**) density in the 2D n^+– p–n^+[p^+] structure for current density $I = 20$ kA/cm^2 [20]

Similar results can be reproduced using a commercial simulator, for the distributed device (Fig. 2.10a) with the following modifications. Two 0.25 μm wide oxide regions have been added to the structure with the silicon width extended to 10 μm (Fig. 2.27a). The current filament state with two filaments is formed at 7 ns (Fig. 2.27b).

2.9.2 Current Filamentation Effect in Double-Avalanche–Injection Conductivity Modulation

The first numerical solution for current filaments due to double-avalanche–injection conductivity modulation has been obtained in [37] to solve the problem of electrical Schottky gate burnout in GaAs MESFET structures. For the current filament solution, a model of the M–i–n^+ structure with a Schottky contact has been used to

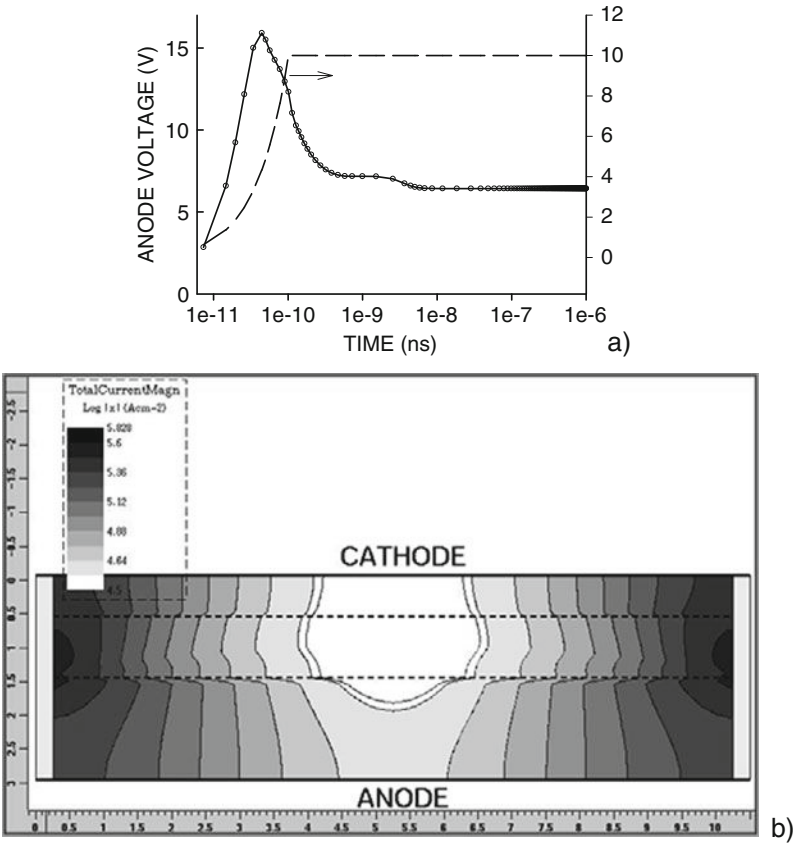


Fig. 2.27 Transient characteristics during the current filamentation effect in the n–p–n structure with a floating p-region at current density 10 mA/μm (**a**) and current density depth profile for the formed state with two filaments (**b**)

obtain the solution for a physical equivalent of the 1D case (quasi-1D case) with a width of $w = 0.2$ μm and the case of filamentation with width $W = 10$ μm. The i-region length of the M–i–n^+ structure is $L_i = 1$ μm, the n^+-region length is $L_n = 0.3$ μm, and the donor concentration in the n^+-region is 10^{18} cm^{-3}.

Conductivity modulation at Schottky diode breakdown is similar to the phenomenon in the p–i–n structure [9].

Two-dimensional distribution of field and carrier density for the filament state is presented in Fig. 2.28. For current filament solutions, typically the electric field and carrier density distribution inside the filament region is close to that of the narrow (1D) structure in high-conductivity state. At the same time, distributions further away from the filament correspond to the state of 1D structure before avalanche multiplication.

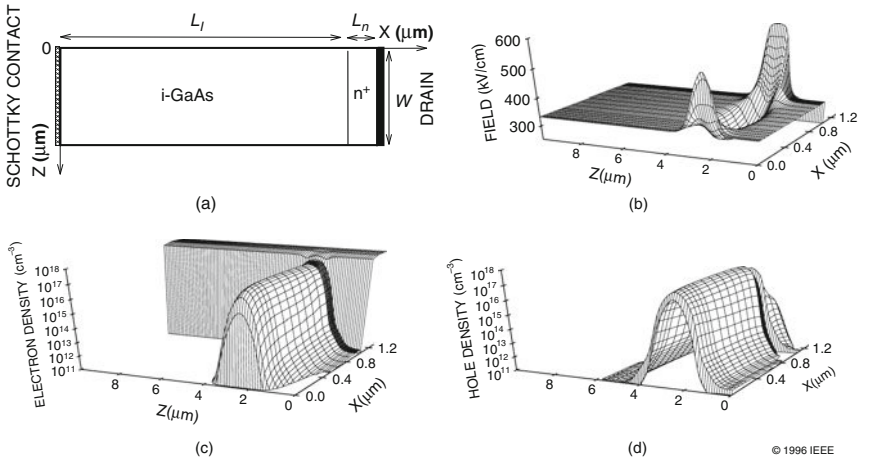


Fig. 2.28 Current filament in10 μm-wide M–i–n^+ structure (**a**) and electric field (**b**), electron (**c**) and hole (**d**) density distribution at 1×10^5 A/cm^2 current density [19]

In case of avalanche–injection, the maximum current density inside the filament is not simply limited below the current saturation in the n^+-region, but exceeds it. This is the result of the current spreading to the n^+-region, which essentially supplies more total current to the filament region than the corresponding saturation current density in the n^+-region. This explains why the very narrow high amplitude filaments, realized in Schottky diode breakdown, are very hard to observe experimentally.

A summary of the simulation results for the silicon p–i–n structure with doping profiles (Fig. 2.5a), expanded total width, and SiO_2 regions at the device boundary is presented in Fig. 2.29 for a current density of 10 mA/μm.

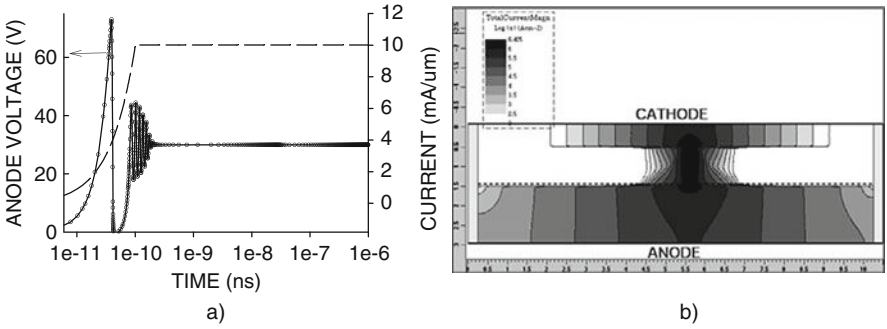


Fig. 2.29 Transient characteristics during the current filamentation effect in the p–i–n structure with a floating p-region at current density of 10 mA/μm (**a**) and obtained final state with two filaments (**b**)

As can be seen, filamentation occurs within a very fast transient process of $t_F <$ 1 ns, providing a rather narrow filament with a width of ~1 μm (Fig. 2.29b).

2.9.3 Current Filamentation Effect in the Case of Double Injection

A similar solution of current filament can be obtained and studied in detail for a p–n–p–n structure with shorted base regions, originally presented in Fig. 2.23a.

To obtain the distributed structure, similar SiO_2 regions have been added and the silicon region width of the structure has been increased up to 10 μm, iterating interdigitated n^+ and p^+ regions for anode, cathode, and blocking junction connections (Fig. 2.30a). The quasi-stationary solution corresponds to the current filament at the right side of the device (Fig. 2.30a), formed during a two-stage process during t_F~0.5 ns [19].

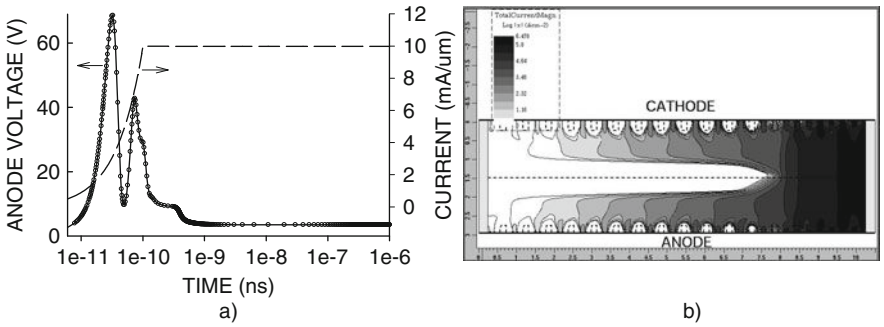


Fig. 2.30 Transient characteristics during the current filamentation effect in the p–n–p–n structure with a floating p-region at a current density of 10 mA/μm (**a**) and obtained final state with two filaments (**b**)

2.10 Summary

In this chapter, major characteristics of the conductivity modulation process have been presented. Timescale of the ESD events allows us to make certain practically important assumptions. Most of the processes in ESD devices can be described in assumption of the adiabatic conditions. Due to the short time of the ESD event, the generated heat is localized within the active area of the devices and typically propagates in the structure on the distance of ~1 μm only. In this case, a rather accurate numerical solution can be obtained using thermal-coupled mixed-mode simulation of ESD device cross sections with relatively small area not exceeding that required for accurate simulation without heating effects.

Due to the same ESD timescale, only electrical mechanisms rather than thermal determine the initial conductivity modulation phenomena and resulting ESD pulse operation transient waveforms.

Thermal and electro-thermal mechanisms [9] become critical only in regimes too close to irreversible failure, thus providing a physical limitation on the maximum pulsed power for ESD devices. In most cases, these phenomena provide constraints for the ESD design that mainly focuses on the provision of appropriate waveforms, dc voltage tolerance, and parasitics reduction. Typically, when the combination of the design targets for ESD device is achieved, the physical limitation of the ESD device operation regime itself occurs beyond the ESD protection window. Therefore, isothermal conductivity modulation mechanisms are directly applicable to the ESD device design, discussed in the following chapter.

In summary, in isothermal pulsed conditions, only four major basic conductivity modulation phenomena are realized in semiconductor structures in addition to the injection in forward-biased diodes:

(i) avalanche breakdown;
(ii) double-avalanche–injection;
(iii) avalanche–injection;
(iv) double injection.

Practically any complex ESD device, as well as standard devices provided by the manufacturing processes technology, or even parasitic devices formed inside integrated circuit, can only provide a condition when either one or a combination of these basic mechanisms is activated.

Often, a more complex domino effect or alternative current path is realized. For example, snapback operation of an SCR device for ESD protection can be initiated first by avalanche breakdown of the blocking junction, then by avalanche–injection in the parasitic NPN structure, and only then double-injection conductivity modulation will take place.

Another example is avalanche–injection in vertical NPN, realized in the BiCMOS process when, depending on the current level, the lateral current conduction can take over the vertical current path. Double-avalanche–injection can be initiated due to n-base conductivity modulation under avalanche breakdown of the p^+–n–n^+ structure.

Nevertheless, in spite of a variety of ESD protection devices and clamps, the physical effects taking place during ESD events are similar to the phenomena in elementary p–n, n–p–n, p–i–n, p–n–p, and p–n–p–n structures discussed in this chapter.

Chapter 3 will demonstrate how understanding of these phenomena can be used to study more sophisticated ESD devices and provide physical limitations of electrical SOA in standard devices. Then, in Chapter 4, clamp-level ESD protection will be discussed using both the equivalent circuits of ESD clamps and the layout topology.

DECIMM^{TM} Simulation Examples for Chapter 2

To download a trial version of the numerical simulation software and request an electronic license key please visit http://www.analogesd.com

To download libraries with simulation examples for this chapter please visit http://www.analogesd.com/Chapter2.html

List of examples is subject to change.

Example 2.1 Numerical analysis of the conductivity modulation in the diode p^+–p–n^+ structure.

Example 2.2 Numerical analysis of the conductivity modulation in the p–i–n structure.

Example 2.3 Numerical analysis of the conductivity modulation in the n^+–n–n^+ diode structure.

Example 2.4 Numerical analysis of the conductivity modulation in the quasi-1D n–p–n structure.

Example 2.5 Numerical analysis of the conductivity modulation in the 2D NPN BJT structure.

Chapter 3
Standard and ESD Devices in Integrated Process Technologies

In Chapter 2, conductivity modulation mechanisms were discussed based upon examples of these mechanisms in corresponding elementary semiconductor structures. The purpose of Chapter 3 is to describe how these conductivity modulation phenomena are realized in both standard devices supported by electrical design rules of given integrated process technology and free ESD devices developed in the process.

The pulsed safe-operating area (SOA) and physical limitations of the operation regime are discussed first for standard devices in typical integrated process technologies. Then, "free" ESD devices are described in greater detail specific to the device type.

The definition of a "free ESD device" assumes no additional process steps in device formation. A free ESD device can be built using only the available mask layers within physical limits of the process. The device can be formed using a self-aligned approach, where variation of the masks will not change the characteristics, or as a non-self-aligned device. In the last case, the device is sensitive to misalignment of the mask layers, which requires the use of additional measures to guarantee repeatable device characteristics (or acceptable parametric yields) within the specified parameter range.

ESD device engineering requires a strong understanding of the integrated process technologies, architecture and process step specifics. Therefore, to support the device-level discussion, Sections 3.1 and 3.2 establish basic definitions. These definitions cover the most typical for integrated analog circuits manufacturing process technology steps and supported devices regions. This material is required for understanding of the subsequent sections. A deeper understanding of the process technology can be found in many published books in the field, for example [39].

The discussion of the pulsed safe-operating area (SOA) in ESD pulse time domain is presented in Section 3.3. Successive sections are focused primarily on the most typical ESD device types.

V.A. Vashchenko, A. Shibkov, *ESD Design for Analog Circuits*,
DOI 10.1007/978-1-4419-6565-3_3, © Springer Science+Business Media, LLC 2010

3.1 ESD Specifics in Integrated Process Technology

3.1.1 Typical DGO CMOS Process with Extended Voltage Components

Complementary MOS (CMOS) processes are the most popular low-cost processes of choice for low-voltage analog design. In the case of the CMOS process, previously developed intellectual property (IP) for digital blocks can be easily integrated with analog blocks.

Taking advantage of accurate mask alignment provided by modern tools, over the recent years these processes were upgraded with extended voltage devices. These extended voltage devices (drain-extended MOS [DeMOS]) enabled the release of analog power products with an operating voltage that significantly exceeds the low-voltage gate oxide limits. For example, creation of non-self-aligned 12 and 20 V NDeMOS and PDeMOS can be accomplished in the 5 V process within the same mask set, while 40 V DeMOS can only be obtained using two additional mask layers for an extended drift region, in order to satisfy hot carrier degradation reliability requirements.

Thus, low-cost CMOS processes enable a combination of previously developed IP for complex digital blocks and lateral high-voltage power devices to support different blocks of analog products: power trains, level shifters, low-dropout (LDO) regulators, and other subcircuits.

Integrated process technology is usually subdivided into front-end and back-end phases.

The front-end phase includes all the process steps (etch, diffusions, depositions) from the initial substrate up to contact formation.

The back-end phase starts from the first metal layer. It includes formation of via's, metal layers and different back-end modules. These modules can include integrated thin film resistors, copper top metals, high-density MIM capacitors, on-chip inductances, laser trimmed fuses, MEMS, and others.

This section presents a simplified description of the front-end CMOS dual-gate oxide (DGO) process flow required to support just six major standard active devices, in combination with ESD specific discussion. The following device formation is discussed:

i) low-voltage (LV) NMOS and PMOS devices for the digital core circuit;
ii) high-voltage (HV) NMOS and PMOS device for interface circuit blocks;
iii) extended voltage (NDeMOS and PDeMOS).

In addition to this, a set of major passive devices (inductors, diffusion resistors, and polycapacitors) is usually formed during the front-end process phase to support the analog product specification.

3.1.1.1 Initial Wafer Material

The process development usually starts with the selection of the initial substrate material. To meet low-cost requirements, the substrate can be non-epitaxial, but in most cases epitaxial substrate material is preferred to guarantee lower latch-up spacing requirements. At the same time, in the case of high-Q inductors implemented on the chip, a very high-resistant substrate is carefully selected to reduce electromagnetic induction current losses.

Typically, p-substrate material is used in extended CMOS processes. Non-epi substrate material is highly resistive. On the contrary, epi substrate material is usually formed on heavily doped p+ substrate with a 2–6 μm lightly doped p-epitaxial layer.

In spite of being more expensive, epi material provides an active silicon layer of a much better quality. From the analog design point of view, epi-material is preferred because of the resulting relaxed latch-up spacing and guard ring design rules, as well as much better sub-block isolation due to the usually grounded p+ substrate.

Respectively, the substrate material has a significant impact on the parameters of the parasitic NPN devices in standard process devices. For example, characteristics of NMOS devices depend on substrate material.

More expensive process technology option is bonded silicon-on-oxide (SOI) wafers. In this case, an active silicon region is separated from the substrate by an oxide layer. The former is isolated by a deep trench region (DTI), thus providing completely isolated devices and circuit blocks. This type of substrate completely eliminates the possibility of latch-up current path between isolated devices.

In high-voltage power devices, the substrate effect can impact the breakdown voltage of the device due to induced charge in the substrate. Usually, it is considered that buried oxide (BOX) provides isolation of ∼100 V per 1 μm thickness.

In the case of the SOI substrate in dc operation regime, the power dissipation limitation should be taken into account, as it provides one of the major limitations of power circuits. However, at least for a relatively thick silicon layer (∼4 μm), the heat generation has a minor impact on operation under ESD pulse conditions, due to a relatively small heat propagation range of ∼1–2 μm during ESD pulse. In this case, elevated lattice temperature is present only in the subsurface area of the device.

Meanwhile, the effects of substrate-generated carriers and the spreading of the injected current into the substrate can significantly change the characteristics of the parasitic devices at ESD conditions or pulsed SOA limits, depending on substrate material. For example, a snapback NMOS device can change characteristics with different substrate resistance, due to a corresponding change of base resistance in the parasitic n–p–n structure responsible for the avalanche–injection process.

Typically, processes steps usually start from precleaning operations for the surface with a high-temperature anneal of excessive oxygen, followed by the formation of an alignment pattern using etch or laser marking. The alignment pattern is used for mask alignment in the following process steps.

3.1.1.2 Device Isolation

The next steps in the CMOS process are intended to form a surface isolation between the devices and device regions on the chip. The regions formed during isolation include a shallow trench isolation (STI) and, if available, deep trench isolation (DTI). Some older or low-cost processes can use junction isolation. Surface isolation is useful because it provides process stability by eliminating hard-to-control surface states and, at the same time, masks shallow implants, creating self-aligned architecture.

An alternative surface isolation in low-cost high-voltage power process is thick field oxide isolation (TFO) or LOCOS. This isolation approach avoids chemical mechanical polishing (CMP) required for STI formation, but provides much bigger minimum isolation dimensions and stress areas with in the so-called "bird's beak" region. In this region, broken bonds of silicon create high-density charge traps. The last significantly impacts opportunities for extended voltage devices, unless special measures are applied to the drift region profile.

STI and TFO can be used in very efficient ways for ESD purposes in order to form free lateral NPN, SCR, and other devices with high-voltage tolerance (Section 3.3). By changing the isolation region dimensions within the physical minimum process dimensions, appropriate voltage tolerance and triggering characteristics can be controlled under appropriate yield.

There are many process variations that form shallow and deep trenches. One of these versions is described below.

An STI isolation formation (Fig. 3.1) starts by Nitride and SiON deposition. These layers are used to mask the silicon surface during etch and high temperature cycles. Then, shallow trench etch removes nitride oxide and silicon at the desired isolation depth (usually at 0.35–0.45 μm), followed by trench liner oxidation. This important step is required to improve the quality of the interface and eliminate interface trap regions that might otherwise significantly impact both the hot carrier degradation (HCD) effects at normal operation conditions and breakdown voltage walkout in extended voltage devices [40].

The following steps include HDP oxide deposition followed by a reverse composite mask and an HDP oxide etch. The next step of isolation formation is chemical mechanical polishing (CMP), which is required to planarise the trench oxide filling regions. Finally, the remaining nitride layer is etched.

Deep trench isolation involves more complex but similar process steps. These steps are partly shared with STI formation. DTI may involve polysilicon trench filing. The DTI in the SOI process is formed to merge with the buried oxide (BOX) region. In high-voltage BCD processes, DTI depth can reach 15–30 μm or even greater.

The remaining open silicon surface outside of the STI and DTI regions we will call composite, referring to the corresponding mask layer as Composite Mask.

Since ESD devices at high current usually support the lateral operation mode, the depth of the trench that defines the current path cannot be neglected.

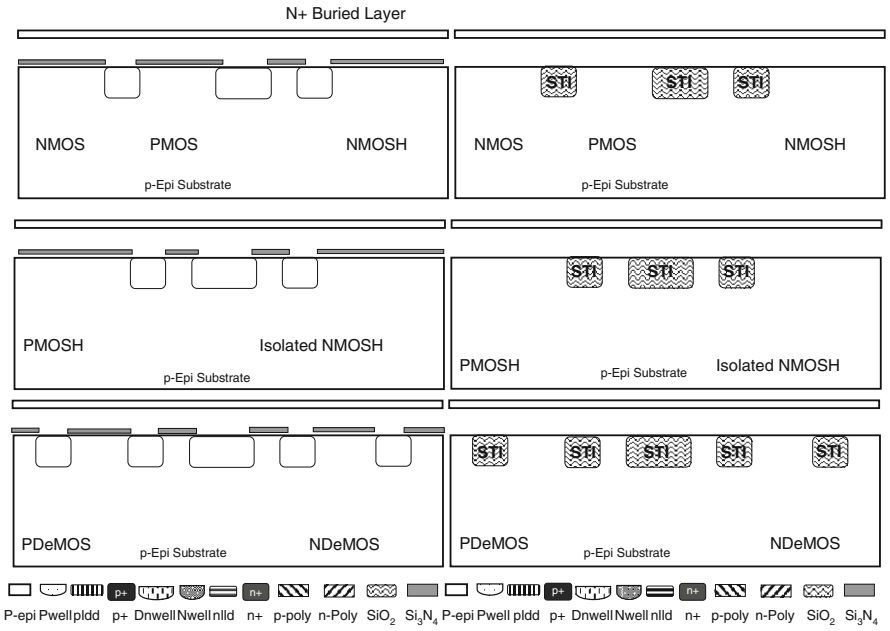


Fig. 3.1 Process flow cross-sections for STI formation

3.1.1.3 Deep Nwell isolation

The following process steps provide the deep N-isolation layer. In this book, this layer will be further called the Deep Nwell (DNWELL) in CMOS processes, or *n*-isolation (NISO) in BiCMOS and BCD processes.

The initial process steps include the pad oxide region etch and growth of the thin, ~10 nm sacrificial oxide layer, followed by a corresponding masking step and a high-energy implant of deep n-type species (Fig. 3.2).

In CMOS processes, DNWELL is used to provide vertical voltage isolation from the p-substrate in the Pwell regions of NMOS, thus enabling the isolated version of these devices. The same layer might be used to implement isolated PLDMOS devices. In BCD processes, the NISO region is a deep region that provides isolation of vertical PNP devices with a P-buried layer (PBL). Both layers are very useful in suppressing noise and coupling between circuit blocks, and in presenting latch-up isolation. Similarly, they are used to provide isolated ESD devices. The isolated ESD devices are used for protection of the voltage nodes relatively to other biased nodes, rather than the grounded substrate node.

In a snapback NMOS ESD device, the DNWELL can be efficiently used to increase the intrinsic base resistance of the parasitic NPN devices in NMOS snapback cells. Availability of the DNWELL implant region enables creation of dual-direction ESD devices (Section 3.3). Without this region, the double blocking

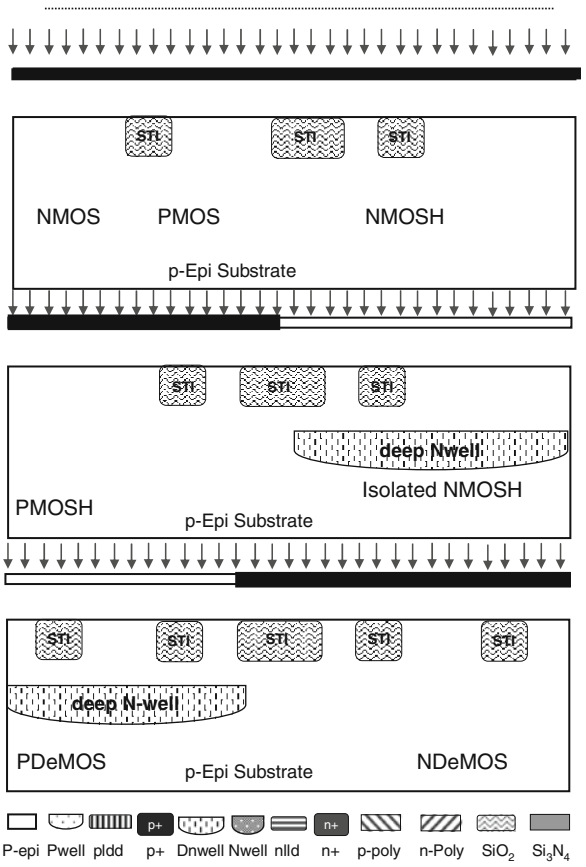


Fig. 3.2 Process flow cross-sections for DNWELL formation

junction regions cannot be realized in the structure. It is assumed that the formation of the Nwell regions at the edge of the DNWELL implant is necessary to complete lateral isolation of the enclosed Pwell regions.

For ESD and latch-up regimes, it should be taken into account that with an added DNWELL region, an additional parasitic vertical PNP device is formed. In this parasitic structure, the isolated Pwell forms an emitter, the DNWELL an n-base, and the grounded p+-substrate a collector. Depending on the thickness of the DNWELL doping level, the gain of this parasitic PNP can be high, and thus cannot be simply neglected. For example, in a dual-direction DIAC ESD device, the additional positive and negative feedback provides very asymmetrical characteristics at positive and negative gate bias (Section 3.3).

A similar p–n–p structure is realized in BCD processes, where the PBL, NISO, and p-substrate act as an emitter, base, and collector, respectively.

3.1.1.4 Well Implants

The most popular CMOS and BiCMOS processes usually include low-voltage and high-voltage CMOS devices. The low voltage of the CMOS provides the advantage of implementation of high-density logic circuit blocks. The high-voltage CMOS devices are used for interface and power pins. Typically, the control and I/O pins support one of the standard voltage levels, for example, 3.3 or 5 V. The voltage level is used to drive the gate of the high-voltage lateral DeMOS or LDMOS devices. Since low- and high-voltage MOS devices require different gate oxide thickness the process is usually supports dual gate oxide (DGO) or triple gate oxide (TGO) in order to allow appropriate scaling of the MOS device gate length.

In low-cost processes, in order to reduce the mask count, Nwell and Pwell regions could be partially shared between LV and HV CMOS devices.

Often, a single drawn mask layer defines both well implants. Usually, for Nwell, this layer is further used to produce an automatically generated Pwell mask. One of the important parameters of high-voltage ESD device design is the spacing between the Nwell and the generated Pwell. Variation of this spacing can provide formation of the graded junction with the desired breakdown voltage level.

For practical ESD device design, it might be rather useful to interfere with the generated layers and perform manual editing of the final generated layers. This approach enables the formation of a variety of new free ESD devices optimized for appropriate ESD protection. The only remaining limitations of such an effort are the physical design rules of the process. Therefore, in-depth knowledge of the particular integrated process technology is necessary.

The most powerful help during design comes from physical process simulations using TCAD tools. In this case, the opportunities of the free ESD devices with altered well implants can be efficiently explored, followed by the experimental test chip verification.

In practical implementation, both the Nwell and Pwell are formed by a combination of several implant steps with different doses and energies (Fig. 3.3). Different species might be used in order to form a complex optimal well profile. The profile may include an anti-punch-through in-depth implant to improve latch-up prevention, a buried channel, or the formation of a so-called empty well. Often, a retrograded well profile is achieved by introducing the LV Nwell implant and the anti-punch-trough (APT) n-type (Phosphorous) implant. During the next masking step, using a Pwell blocking mask, both the LV Pwell and the anti-punch-through p-APT implant set are applied using different dose and energy p-type (Boron) species.

Depending on the process guidelines, one or two additional masking steps are done to create a threshold voltage implant for PMOS VTP n-type (Arsenic) and a threshold voltage implant for NMOS VTN p-type (Boron). These shallow implants form the upper portion of the wells. Thus, wells are formed in three regions: VT, well, and APT.

When low-voltage wells are formed, similar steps are completed for high-voltage wells HNWELL and HPwell.

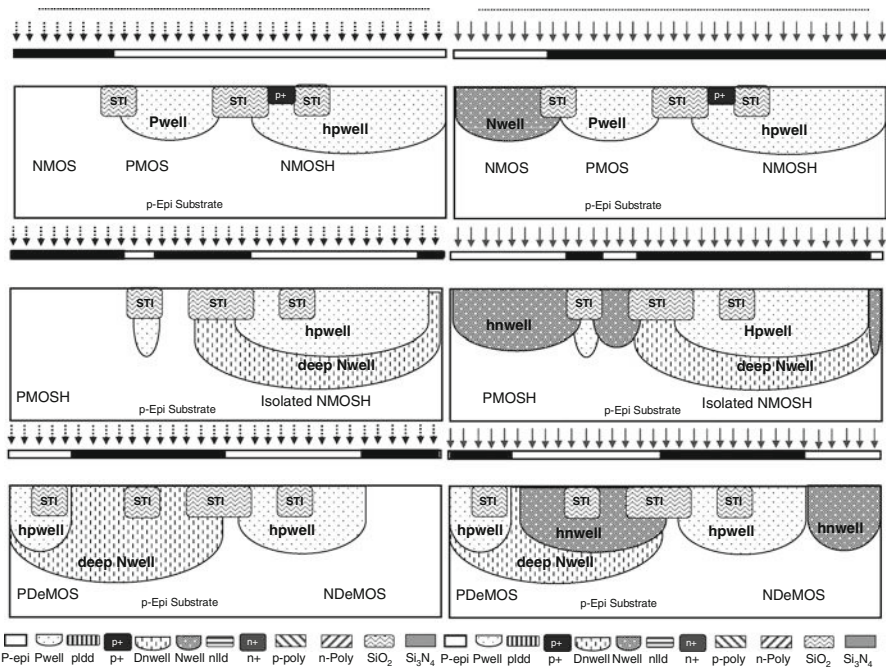


Fig. 3.3 Process flow cross-sections for Pwell and Nwell implant

3.1.1.5 Gate Oxide

The next typical steps form the thin gate oxide layer for the LV CMOS device and thick gate oxide for HV CMOS devices.

At first, the sacrificial oxide is removed, followed by growth and etch of the recovery oxide. These steps are required to eliminate defects formed on the silicon surface during well implant steps. After subsequent cleaning, HV gate oxide thermal oxidation is first used to form the thick gate oxide.

In 3.3 V high-voltage CMOS devices, the oxide thickness is ~6.8 nm. Then, a gate oxide mask is developed, followed by gate oxide cleaning in the places of future LV CMOS devices.

Thin LV Gate oxide formation is achieved by corresponding thermal oxidation. A typical thickness of the gate oxide for 1.2 V devices in 0.13 μm processes is ~2 nm. During LV gate oxide growth, the HV gate oxide continues to grow up to the final thickness of 7 nm (Fig. 3.4).

3.1.1.6 Polygate

As soon as thermal oxidation is completed, polysilicon (poly) deposition is also fulfilled, followed by Poly mask and etch. The etch stops on gate oxides for both

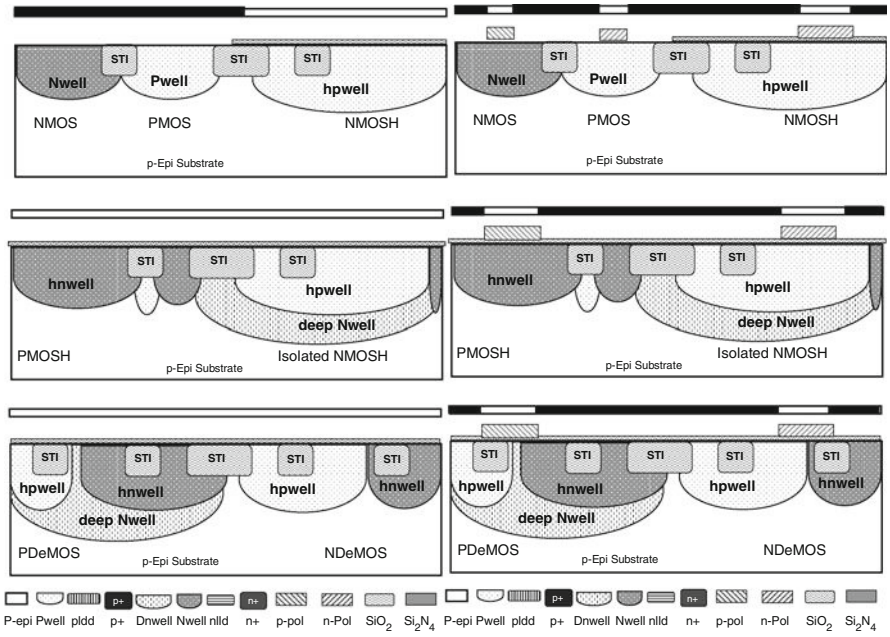


Fig. 3.4 Process flow cross-sections for gate thin oxide formation and poly deposition

LV and HV devices. The minimum length of the LV gate is the most challenging optimization parameter to meet.

For ESD purposes, the devices can be produced using minimum LV gate length in conjunction with HV poly regions. This enables creation of split gate structures, gated diodes, and a control electrode in other ESD devices. However, the voltage tolerance of such devices both to dc voltage levels and to long-term reliability operation requires additional assessment.

In free ESD devices with MOS regions, the gate oxide absolute maximum rating is the major limiting factor for voltage tolerance. For example, a 1.2 V CMOS device might provide a perfect snapback characteristic for 2.5 or 3 V pins. However, long-term reliability concerns will not allow application of such a clamp in the product due to gate oxide overstress.

Another ESD aspect related to gate oxide quality is robustness in short-pulse electrical overstress. Depending on the quality of the dielectric and the process itself, the maximum pulsed gate voltage before burnout may vary significantly. Usually, in short pulse conditions, the critical voltage for gate oxide obeys the power law [41].

Recently, more complex gate materials with a high-K dielectric and sophisticated metal gate systems specific to NMOS and PMOS devices have been commercialized. However, application of these processes so far is confined mostly to digital products.

3.1.1.7 Lightly Doped Drain Implants

The scaling of the CMOS devices below 0.5 μm processes requires lightly doped regions. The regions provide robust hot carrier degradation performance, allow a high breakdown voltage, and eliminate gate induced drain leakage (GIDL) and drain induced drain leakage (DIDL).

Before n- and p-lightly doped drain (NLDD and PLDD) implants are performed, an oxide etch and clean from all silicon and poly areas is necessary, followed by the application of the rapid thermal process (RTP) to seal poly and clean, creating a few-nanometers-thick oxide on silicon in the place of all subsequent implants (Fig. 3.5).

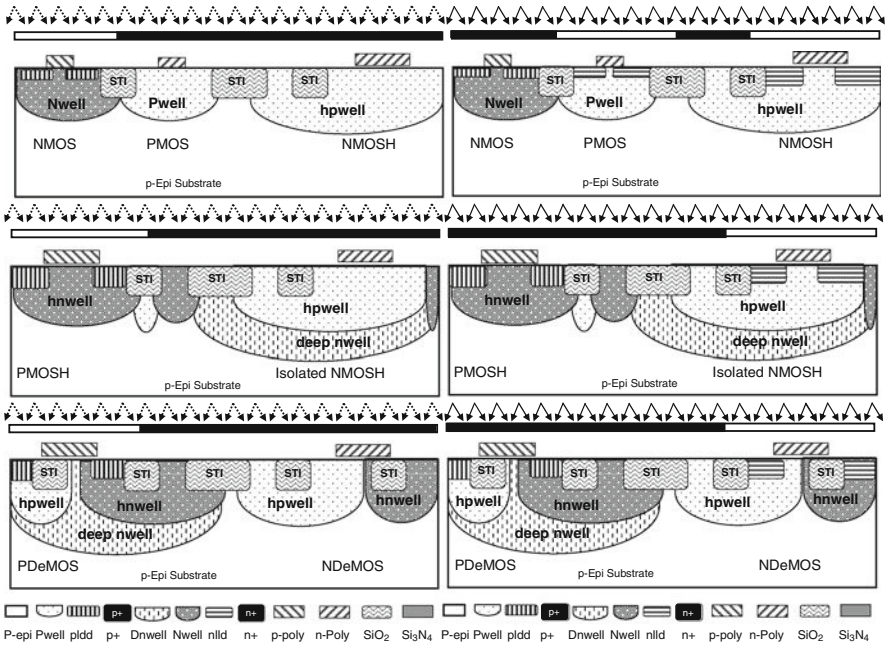


Fig. 3.5 Process flow cross-sections for pldd and nldd implant

For all LDD implants, the poly gate region acts as a masking layer that creates self-aligned conditions.

In submicron CMOS process, an LV NLDD n-type is formed with arsenic, followed by a Halo NLDD implant in the same opening. The halo implant in the LV NMOS device is a p-type and made completely of BF_2. The HV device may receive a separate HNLDD n-type implant (phosphorus).

Then, an LV PMOS PLDD mask p (BF_2) and halo implant n-type (As) are performed, followed by an HV HPLDD mask BF_2 implant for the HV PMOS device.

Both PLDD and NLDD lightly doped drain implants might be very useful in creating ESD devices with a triggering voltage that corresponds to either n+-PLDD or p+-NLDD.

Implementation of the free self-aligned avalanche diodes achieved by overlapping the corresponding LDD mask layer with an n+ or p+ layer is very advantageous because it forms a lateral surface structure. In 0.5 μm CMOS processes, the breakdown voltage of the devices is ∼6–8 V and therefore often suitable for the creation of the 5 V avalanche diode and avalanche diode referenced snapback clamps. These devices provide appropriate targeting of the ESD protection window by providing a much better alternative to the n+ -Pwell and p+-Nwell blocking junctions with a breakdown voltage of ∼11–14 V.

Creative manipulation of different LDD combinations may result in the creation of other surface blocking junctions and surface BJT ESD devices with low-voltage triggering characteristics.

The major limiting factor for a surface junction with an LDD region is the formation of surface states that may create unstable conditions for breakdown voltage and enable walkout and abnormal leakage variation. Therefore, such free device design requires further qualification that involves reliability testing and understanding of the yield impact, especially if non-self-aligned layers are involved.

Another ESD application of the NLDD regions in ESD design is the formation of surface saturation resistors.

3.1.1.8 Spacer Formation and NPLUS and PPLUS Implants

The following steps create the spacer region. Together with the POLY region, the spacer provides an additional self-aligned extension of the polygate to mask NPLUS and PPLUS low-energy implants at a certain distance from polygate edge. This allows implants to be shifted some distance apart from the gate edge, thus forming a desirable implant profile for electric field distribution, which reduces the hot carrier degradation effect and increases breakdown voltage of the device.

There are many different recipes for spacer formation that depend on the tools involved and the process architecture. One of the examples of spacer formation is presented below. It includes TEOS oxide liner deposition forming a ∼ 20 nm thermal oxide, followed by a ∼40 nm nitride deposition and addition of a 120 nm thermal. Then, an anisotropic plasma etch of the oxide results in spacer formation, followed by a nitride spacer chemical etch step (Fig. 3.6).

Thus, together with the poly region, the spacer masks subsequent NPLUS and PPLUS implants.

After an NMOS source and drain mask, an n-type (arsenic) implant and graded junction n-type implant (phosphorus) are made, followed by anneal. Then, the corresponding PMOS source and drain mask and p-type (boron) implants form the graded junctions.

Corresponding PPLUS and NPLUS implants are made to form well taps in the CMOS devices.

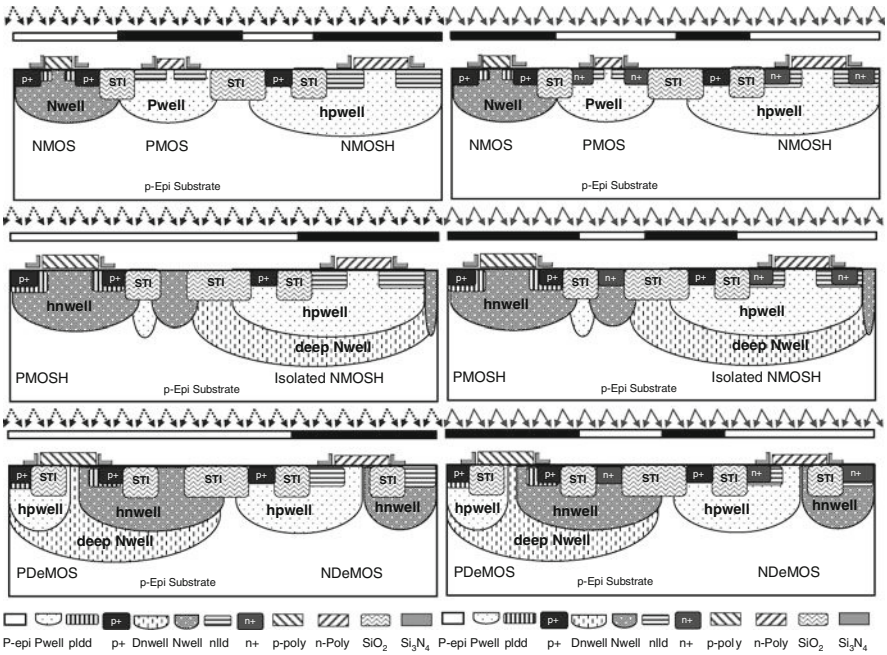


Fig. 3.6 Process flow cross-sections for spacer formation

The third spacer etch step consists of a dry oxide etch, which removes the spacer footer oxide from the top of the poly and silicon, followed by cap oxide deposition.

3.1.1.9 Activation and Silicidation

Rapid thermal process anneal (RTP) is used to activate the completed shallow implants. Then, a cap oxide is etched. After the etch, only the L-shape spacer is left at the polygate edge of the CMOS devices (Fig. 3.7).

The silicide exclusion mask, SALEX, is developed to mask the area where silicidation should be avoided. In standard devices, silicidation exclusion is used for diffusion and poly resistors.

In ESD devices, the silicide exclusion region is absolutely necessary in order to provide a drain ballasting region in snapback NMOS devices and a floating drain in SCR devices.

In the case of cobalt material, the silicide is formed using cobalt precleaning for the silicon and polysilicon surface in order to remove oxide. Then, cobalt deposition is made (∼10 nm Co). Other additional layers, for example TiN, can be used as well. After the first step of silicidation, RTP is completed. The cobalt that did not react with silicon is removed, followed by the second silicidation anneal RTP step.

There are two important considerations for the silicidation process in ESD devices.

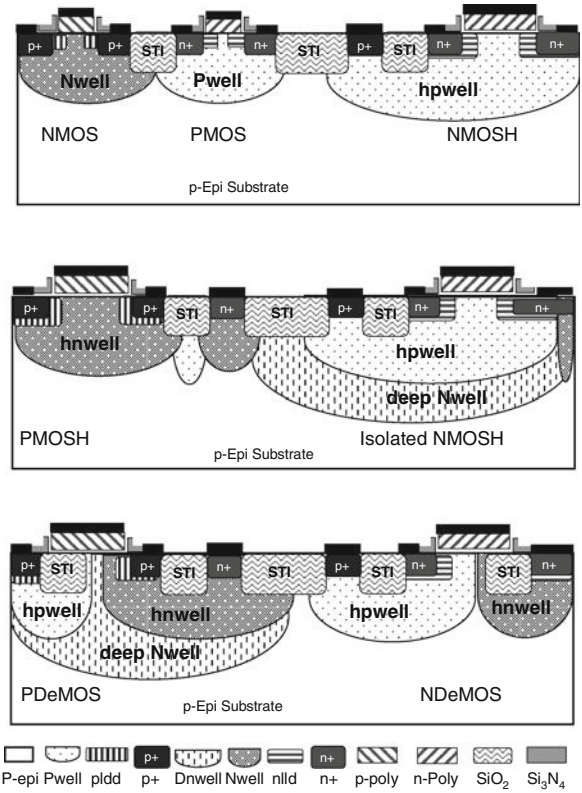


Fig. 3.7 Process flow cross-sections for activation and silicidation

First of all, positioning of the silicide exclusion masking layer at the drain side is critical for the performance of snapback ESD devices.

If the mask is aligned with the gate edge, then there is a risk of formation of a very narrow silicide "sliver" region at the gate edge can result in low performance of snapback NMOS devices, due to melting of the silicide at lower temperatures during ESD stress and penetration in the active region of the device.

Such a narrow residual silicide region formed along the gate impacts local current ballasting. Therefore, the maximum ESD current supported by such a device is significantly lower. To overcome this issue, it is necessary to draw the silicide exclusion mask layer overlapping the gate either completely or to the middle of the gate. In ESD clamps with a grounded gate, this measure usually provides no disadvantage. However, an unsilicided gate region might be undesirable in high-speed I/O where ESD and I/O functionality is shared within the same device structure. The second option might be limited by the alignment tolerance.

Another ESD consideration for the silicidation process is related to the formation of minimum composite regions as a part of the ESD devices, for example in ESD diodes. In this case, artifacts from the nitride and oxide etches may reduce the actual

opening of the contact diffusion surface, thus significantly increasing the resistance. To overcome this issue, ESD diodes might be optimized with a diffusion region size larger than the minimum dimensions of the process.

3.1.1.10 Contacts and Backend

The contact formation steps include a self-aligned passivation layer deposition, a pre-CMP deposition of 50 nm nitride, planarization by CMP, a contact mask, and a contact etch (Fig. 3.8).

This is the last step of the front-end portion of the process flow.

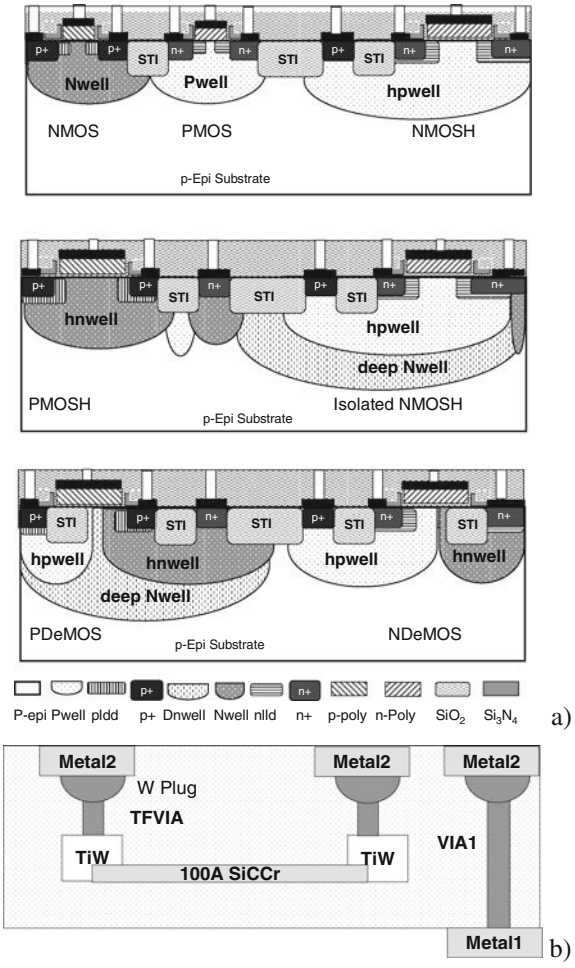


Fig. 3.8 Process flow cross-section for contact formation (**a**) and thin film resistor (**b**)

At the back-end, the ESD current density should be taken into account in order to follow certain process-specific multiplication rules for the electromigration limits.

In the case of some processes, there exist limitations on stacking the vias and contacts or vias on top of vias. ESD clamp design with backend consideration is discussed in Chapter 4.

Another ESD aspect is the possibility of formation of ESD devices under the pad. This usually requires a minimum number of metal layers on top of the device.

3.1.2 ESD Specific for BCD and BiCMOS Integrated Process Flow

3.1.2.1 Generic Process Flow

Similar cross-sections can be drawn for the BiCMOS and BCD process flow. To reduce the volume of information presented in this chapter, these process flow details will not be presented. However, most of these process steps can be understood using the typical BCD devices' cross-sections presented in Section 3.2.3, with the combination of the brief discussion presented in this section.

From a cost perspective, BCD and BiCMOS processes provide a significant variety of different recipes, due to the effort to combine process performance with optimized cost. Therefore, a typical way to integrate such processes is to share mask layers steps between CMOS, BJT, and DMOS devices.

Another practical scenario that targets low cost and requires customized ESD devices is the case of process application to the products that do not require all mask steps. For example, some analog products in DGO BiCMOS process can be implemented without process steps that create a low-voltage CMOS part, which therefore can be removed from the process flow.

A somewhat generic integrated process BiCMOS flow is presented in Fig. 3.9. In real cases, some steps can be different, depending on the specifications of the required set of integrated devices and their parameters.

The starting epi-material usually has an orientation of <110>. Normally, the P-type substrate is used with grown P-epi or N-epi. Alternative substrate can be implemented using the silicon on oxide (SOI) process with a bonded wafer. For 0.5 μm analog processes, the typical epi thickness is 2–4 μm and the buried oxide (BOX) thickness is ~1–2 μm.

The BiCMOS process adds at least one vertical NPN BJT to the CMOS process. In the BCD process, usually BJT, CMOS, and double-diffusion MOS (DMOS) devices are formed.

Substrate selection for the process depends on the voltage and isolation spec. Unlike in the CMOS process, one or more epi growth steps are required to form the subcollector and base regions of the bipolar devices.

A typical set of available devices in the BCD process for high-voltage (HV) power application includes power high-voltage devices NLDMOS, PLDMOS, PMOS; complementary NPN and PNP devices; 0.5 μm 5 V CMOS devices; core

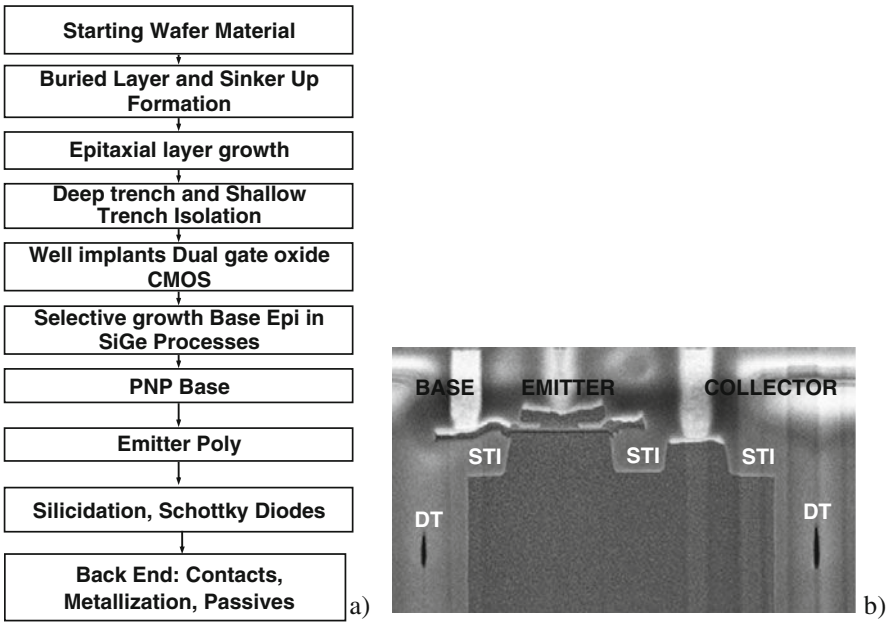


Fig. 3.9 Generic simplified integrated BiCMOS process flow (**a**) and cross-section for 0.24 μm Si–Ge NPN BJT (**b**)

logic low voltage (LV) 0.5–0.13 μm CMOS; different HV and LV tolerant diodes; optional thin film resistor (TFR), HV-isolated capacitor.

All devices must be substrate isolated above the high-voltage (HV) level on additional LV power supply level to enable boost circuit driver operation. A typical advanced process for BCD process flow is presented in Fig. 3.10.

3.1.2.2 Subcollector and Substrate Isolation Regions

In order to support vertical BJT devices in the process and provide deep junction isolation, there are several deep layers that are generated in the process.

Typically, a heavily doped N-type buried layer (NBL) and P-type buried layer (PBL) are implanted prior to epi growth in the high-voltage process.

If the process needs to support high voltage, then the N-epi region thickness is also high. In this case, to contact the buried layer, a stack of sinker layers is required. For example, two sinker implants, NSINKERUP and PSINKERUP, can be made prior to the epi growth or in the intermediate steps of the growth. Then, after the epitaxial layer is fully grown, additional high-dose high-energy implants, NSINKERDOWN and PSINKERDOWN, are made, in order to connect the buried layers and the deep sinker region to the surface of the device.

To isolate the P-collector region in a PNP BJT device, an additional NISO region is used, with the help of a high-energy implant. This implant completely isolates the

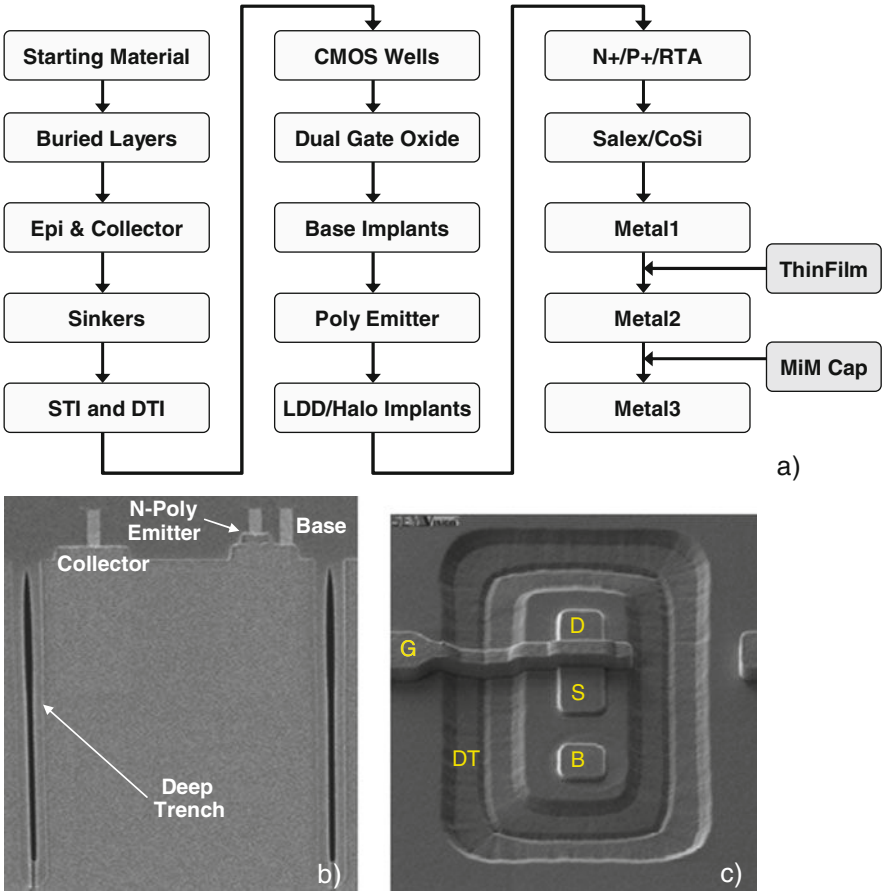


Fig. 3.10 Example of process flow for BCD process (**a**) example cross-section of 20 V NPN structure with deep trench isolation (**b**), and top view of deep-trench (DT) isolated 0.25 µm CMOS (**c**)

PBL vertically. The NBL and NSINKER regions can perform lateral-junction-type isolation. Alternatively, lateral isolation can be completed via the Deep Trench that is available in the process.

An intelligent use of the buried layer and sinker implant is very productive in implementing vertical and lateral self-aligned blocking junctions and avalanche diodes suitable for ESD purposes. The vertical devices not only eliminate the effects of the surface states but also provide advantageous in-depth heat dissipation in the silicon region.

After completion of the N-epi growth, surface sinker regions are formed, followed by an additional thermal anneal that results in diffusion of the implants in all directions. The thermal processes provide not only a significant vertical diffusion of the implanted species, but lateral diffusion as well. Therefore, for free ESD devices

with subcollector regions, it is important to take into account that the real boundaries of the diffused regions are significantly deviated from the originally drawn mask layers.

Highly diffused NISO, NBL, and PBL implants can be patterned with the minimum mask dimensions and thus control gain of the parasitic BJT device structures and the breakdown voltage [42]. The last can be used as a reference for the triggering voltage. The NSINKER can be used for current ballasting in ESD structures.

Depending on the actual process doping profiles and device spec, buried avalanche ESD diodes can be formed using PSINKER-NBL or NSINKER-PBL junctions.

The subcollector region provides for many possibilities of implementing different SCR-type structures with in-depth current conduction, which results in withstanding of much higher ESD power, due to heat dissipation in the bulk silicon material.

3.1.2.3 Isolation BCD Process Steps

Next step is similar to the CMOS process: the formed surface isolation can be a thick field oxide or a shallow trench (STI), or both STI and DTI. Thus, the misalignment of DTI and the deep layers must be accounted for.

Appropriate device isolation is a rather critical factor of the BCD process. At the applied high voltage of ~100 V to the regions on the top of the substrate, the injected carriers can travel the distance of 300–700 μm. This effect can potentially initiate such undesirable effects as latch-up, cross-talk, and interference between the ESD devices and active circuitry, for example, with the switching power LDMOS array.

In the advanced BCD process, the trench, with a depth of ~10–20 μm, is combined with the buried and isolation regions. In a high-voltage process of 50–100 V, an additional P-isolation region (PISO) is formed to connect the p-substrate, thus providing high-voltage junction isolation in a thick epi-layer.

3.1.2.4 Collector and Initial CMOS Regions

Usually, the epitaxial region is of N-type and thus immediately provides n-collector regions for vertical NPN devices.

For vertical PNP devices, the N-epi region is overcompensated by the P-collector implant.

Taking into account the thermal budget, the CMOS module mask steps are performed next, up to the level of well implants and dual gate oxide processing. The major thermal budget contributors include the high temperature rapid thermal anneal (RTA) required to form corresponding layers and regions.

Then, for example, in the SiGe process, selective base epi-growth is made, followed by the formation of the polyemitter, remaining CMOS regions, shallow implants and activation. Alternatively, the shallow internal and external base region implants are made, followed by the formation of the polyemitter and polygates with spacer processing.

The following steps include shallow implants used to form device contact regions. Finally, different front-end modules are fabricated to support back-end based passive elements.

The BiCMOS process provides lots of possibilities of engineering free ESD devices. Required parameters within the ESD protection window range can be achieved using a combination of different implanted regions in order to form parasitic n–p–n, p–n–p, and p–n–p–n structures and appropriate blocking junctions.

Dual-direction devices based upon BJT base-emitter regions will be demonstrated below. The presence of the subcollector region also allows the forming of bipolar SCR devices (BSCR) [7].

Combination of the CMOS implants with the BJT part of the process might be very useful as well. For example, the Nwell can be used to reduce the breakdown and voltage of the standard NPN devices inside the ESD protection window.

The cross-section of the most typical CMOS and BCD devices, including parasitic structures responsible for save-operating area limitations, will be discussed in the next section, thus proving additional insight into the BCD process flow.

3.2 Safe Operating Area in ESD Pulse Regime

Understanding of the pulsed SOA in ESD time domain is critical for successful ESD design. Namely, control of the electrode-dependent voltage range between maximum operation regime and SOA determines the so-called ESD protection window. In a simplified way, a certain voltage range is customary for ESD protection. The range is confined between the absolute maximum operation voltage level of the analog circuit pin at normal operation conditions and real absolute maximum for the pins in the given ESD pulse regime. The last is practically determined by the pulsed absolute maximum voltage of devices connected to the pin within specific ESD test conditions. The test conditions may provide different coupling to the control electrodes of the devices connected to the pins.

The definition of the absolute maximum limits assumes that the reversible operation of the devices is not guaranteed if the device operating regime exceeds the limits. The product might be expected to fail due to electrical overstress (EOS) if the absolute maximum limits for current and voltage at the pins are exceeded. Respectively, in the ESD condition, some devices connected to the pin are expected to fail if the pulsed voltage is over the limit, thus resulting in ESD failure.

Thus, if the ESD protection window is well understood, an ESD solution can be engineered by implementation of the voltage limitation of the ESD device consistent with the window.

In reality, understanding of the ESD protection window is much more complex. For ESD pulse conditions, both lower and upper limits of the window are strong functions of stress conditions. For example, ESD protection windows for the CDM and HBM pulses might be totally different. This effect is related to the delay times in coupling of the control electrodes and the time of the current transient itself.

Thus, the ESD protection window should be considered in the time domain as signal waveforms that include many factors, even the history of previous ESD events. Another critical aspect may be related to the fact that the ESD protection window for the same pins can be very small and even negative depending on the conditions of the internal circuit devices.

In general, the absolute maximum voltage of the supported device in ESD pulse regime is limited by current instability phenomena, and the maximum level depends on the control electrode conditions that control current incoming into the avalanche multiplication area.

The conditions of the control electrode itself are transient in nature and dependent on the internal circuit and ESD pulse waveform. They are often hard to identify even when using ESD compact model analysis.

The next section highlights one of the methodologies adopted to overcome the difficulties in defining the ESD protection window. At least for devices interfacing with external pins, ESD pulsed safe operating area (SOA) can be and should be well characterized in a wide range of control electrode parameters. This helps gain at least a starting level of confidence in finding the correct ESD protection window.

Two methodologies measuring pulsed SOA in the ESD time domain are discussed below, after a description of the conventional understanding of the SOA in the reliability field.

3.2.1 SOA and Current Instability Boundary in Reliability

To guarantee reliable long-term semiconductor device operation, a number of constraints are usually imposed on operation regime.

In transistor structure, these constraints are dependent upon the electrical regime determined by the applied voltage or current to the control electrode.

Usually more or less complete information about maximum and absolute maximum ratings for devices, as a function of their parameters and operation conditions, is provided in electrical design rules (EDR) for integrated components, or in the datasheet for discrete components. For example, the information can provide the expected operation currents $I_{\max}$, voltage $U_{\max}$, dissipated power $P_{\max}$, and other parameters specified for the given frequency, time domain, ambient temperature, and other conditions.

The long-term reliability parameters of the device are guaranteed only within maximum rating limits.

In the range between maximum and absolute maximum ratings, only the short-term reversible operation of the device is guaranteed. Above the absolute maximum rating, the device survivability is no longer guaranteed and the device may fail irreversibly immediately after a limit violation in any time domain.

When an analog product is designed, the maximum and absolute maximum ratings specified for each pin or pin groups, depending on the circuit design specification. In this case, the internal devices are used with appropriate margins that

may not even be related to the absolute maximum limits, but are limited by circuit performance: noise, linearity, functionally or simply a product spec that may be way below the process limitation. Therefore, usually a certain margin in the voltage domain can be expected when no ESD devices are connected. In real application, these ESD devices might be the analog circuit components that provide physical limitation of the absolute maximum voltage for the given analog circuit pins.

The reversible area of operation is represented by an area under a curve drawn in *I–U* coordinates. This area is called the safe-operating area (SOA) for some given conditions, for example, room temperature conditions.

Usually, SOA is plotted for guaranteed maximum ratings in order to provide the customer with an SOA that is reliable for long-term operation. Similar plots can be constructed for absolute maximum ratings and for different pulse and temperature conditions.

The boundary of this SOA is determined by the limitations of the device in electrical and thermal regimes. These limitations are determined by a broad spectrum of electro-thermal physical effects.

The SOA for maximum ratings is limited in most cases by long-term reliability parameters. Power dissipation may also provide the major limiting factor in a typical device's SOA. These factors are related to different degradation phenomena, for example, mean time before failure (mtbf).

On the contrary, the SOA for absolute maximum ratings is based upon physical limiting factors of a more rapid and instantaneous nature. These factors are related to different electrothermal current instabilities that are realized in the conditions of corresponding electro-thermal conductivity modulation effects [9].

In a simplified case, the SOA of output characteristics of an NMOS power device at positive bias (Fig. 3.11a) is limited by three conditions: $I \leq I_{max}$; $U \leq U_{max}$; $I \cdot U \leq P_{max}$.

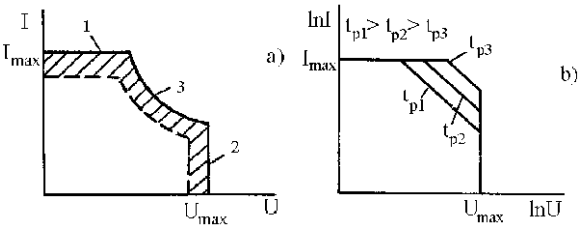


Fig. 3.11 SOA for dc (**a**) and pulsed (**b**) regimes: (1) $I=I_{max}$, (2) $U=U_{max}$, (3) $P=P_{max}$ (where t_P is the power pulse duration)

The current limitation $I \leq I_{max}$ is usually set by either backend electromigration limits or the maximum saturation current that can be obtained from a device at corresponding maximum gate bias (Fig. 3.11a, region 1).

On the contrary, in low power conditions, the voltage limitation $U \leq U_{max}$ is usually determined by the isothermal avalanche breakdown and conductivity modulation effects (Fig. 3.11a, region 2).

The third major limiting factor is the dissipated power limitation $I \cdot U \leq P_{\max}$. Electrical, electro-thermal, or thermal instability effects initiated by Joule heating usually determine this limitation. In dc operation, this limitation depends upon the ability of the device to provide heat dissipation, for example, the thickness of the semiconductor material and the efficiency of the heat sink.

A different scenario is introduced in pulsed operation regime. At shorter pulse or a reduced duty factor, the thermal effects are reduced, providing a dominant role for electrical instabilities [9]. In this case, the final irreversible breakdown or burnout of the device is still related to local melting. However, the root cause of the local melting is the electrically formed current filament state that subsequently provides localized heat generation of such tremendous amplitude that it is capable of local melting in the device, even in short pulse regime [9, 43–45].

The main concept of SOA makes sense if there is an operation regime with a certain amount of time before failure. Apparently, SOA for device operation in dc operation regime with, for example, $t_{\text{mtbf}} \sim 10^5$ h can be significantly different from the SOA of the same device in pulsed regime, for example, with a pulse duration of $\sim$1 μs and low 0.1% duty factor, even if the same current–voltage regime is realized. Then, SOA in a single nanosecond pulse domain can be significantly different from that in a microsecond time domain. Thus, depending upon the regime, different SOA can be constructed.

Furthermore, since SOA depends on the given operation regime, SOA extension could be expected from most cases of device operation in the pulsed regime. This is observed due to suppression of heat dissipation limitations (Fig. 3.11b). However, as it will be shown below, this not always true in ESD pulse, due to electrical instability and conductivity modulation that might be initialized by the dV/dt effect, resulting in significant displacement current in the device junction or control electrode coupling.

3.2.2 Pulsed SOA for ESD Regimes

For normal operation regimes, long-term reliability parameters can be calculated using accelerated test results. The physical mechanisms are related to different degradation effects in the semiconductor structure and metallization; hot carrier degradation, negative bias temperature instability (NBTI), electromigration, mutual diffusion, and many others.

Such SOA limitation usually occurs due to various effects acting in different time domains outside of the ESD pulse domain. Even in the case of electrothermal instabilities, the typical time is comparable with the thermal heating time constant in the device structure which falls in the microsecond time domain.

On the contrary, in the ESD test, the device is expected to withstand a limited count of pulses in the 1–100 ns time domain. Moreover, after ESD test evaluation, the components are no longer guaranteed for long-term operation parameters and not shipped to the customers.

The goal of ESD SOA and instability boundary measurements is usually two-fold: (i) evaluation of pulsed SOA for standard devices and (ii) measurement of the instability boundary for ESD devices to help attain appropriate ESD clamp design.

3.2.2.1 Standard Devices

In standard package level spec ESD pulses (see Introduction), the rise time is usually within 0.1–10 ns and the pulse duration is in the time domain of ~10–100 ns. In addition, a significant time, ~1 s, is provided between repetitive ESD zaps. Thus, ESD stress can be considered as a single pulse event in terms of heat dissipation in spite of the fact that some charge can be injected into the dielectric or interface regions. Although, a more complex cumulative damage scenario can also be observed (Chapter 4).

Thus, pulsed SOA in the ESD time domain (further as ESD SOA) is a SOA measured for specific pulse conditions. In principle, this SOA depends on the specific pulse waveform and is expected to be different for HBM, MM, and CDM, as well as IEC and CDE system level test pulses.

The most common characterization technique for ESD SOA evaluation combines TLP measurements with constant gate bias. This technique is quite informative and relatively widely available.

However, the level of confidence in these measurements is impacted by the correlation between TLP and real ESD pulse conditions. A helpful consideration is that the same TLP measurements are usually used for both ESD device evaluation and pulsed SOA. Thus, the methodology provides rather informative practical results, as demonstrated below.

The test setup for pulsed SOA evaluation using a standard TLP system is presented in Fig. 3.12. This technique applies TLP pulses to the device under test (DUT), for example, NMOS device under constant gate bias on the gate electrode (Fig. 3.12). The setup includes a power source to provide gate-source bias, base-emitter bias, or current. Otherwise measurements are similar to conventional TLP.

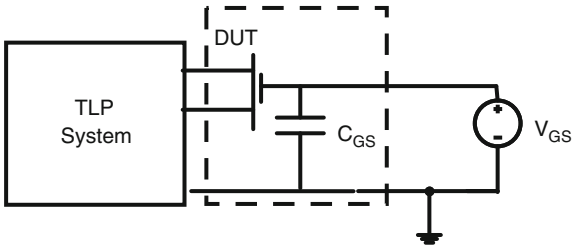


Fig. 3.12 Setup for pulsed SOA evaluation using TLP measurements

BJT devices can be evaluated similarly with the initial conditions of first voltage and then initial current for the base electrode. This mixed regime complicates a possible extraction of the current gain, due to an unknown base current realized in

pulsed mode. However, as it will be demonstrated below, even in these conditions, the pulse SOA can still be obtained, therefore helping the ESD design.

There are several limitations for this setup. First of all, only certain size devices can be characterized, due to a limitation of TLP accuracy. Usually, these devices are not less than 10 μm in total width, at the same time TLP voltage amplitude is usually limited to ~500 V. This limits both characterization of the power arrays at a current below 10 Apms, and the ultra high-voltage device (UHV) within an operating voltage above 500 V.

The system has internal load characteristics of 50 Ω due to corresponding coaxial cables used in its design.

Another problem of TLP SOA characterization is related to the control electrode coupling. Usually, average TLP proportional voltage and current signals are extracted from the TLP waveform in a range between 70 and 90 ns for the standard 100 ns pulse. However, in a relatively big device, for example, the NMOS array parasitic gate-drain oupling capacitance is large enough to influence the test results.

Since the external gate bias voltage source is connected through the resistive (~0.1–1 Ω) wires with a parasitic inductance of ~10 nH, the assumption of constant gate bias during the TLP test is generally false. Before measurement in the 70–90 ns time domain, the device can turn on in snapback. This results in an incorrectly low measured critical voltage in comparison with true value.

To overcome this issue, it is necessary to prefabricate an additional gate-source capacitor on the chip in order to hold the gate bias constant. All SOA presented in this chapter are collected for a device with such decoupling (with an approximately 1 mm total width). The size of the capacitor can be obtained from a simple compact model simulation. In CMOS devices, this size depends upon the ratio between the parasitic gate-drain C_{GD} and gate-source C_{GS} capacitance in the device. In high-voltage devices with long drift regions, the required on-chip capacitor is rather large and comparable with the device size, if it is desired to keep gate voltage coupling within half a volt.

3.2.3 ESD SOA for Typical Devices in BCD Process

Supported devices for analog circuits can be characterized using a pulsed SOA approach.

The case of the 100 V BCD process is presented in Fig. 3.13, together with a pulsed SOA measured using the TLP technique.

The exact set of the devices depends on the process spec. Examples of the process in this section include the 7 V NMOS and PMOS device, the *N*-channel lateral DMOS (NLDMOS) and PLDMOS devices, as well as the 20 V vertical NPN and PNP devices.

In the NMOS (Fig. 3.13a) devices, a parasitic NPN is formed by the n^+-drain acting as a collector, the n^+-source as an emitter, and the Pwell region with p^+-tap diffusion as a p-base. Similarly, BJT regions can be identified in P-type devices.

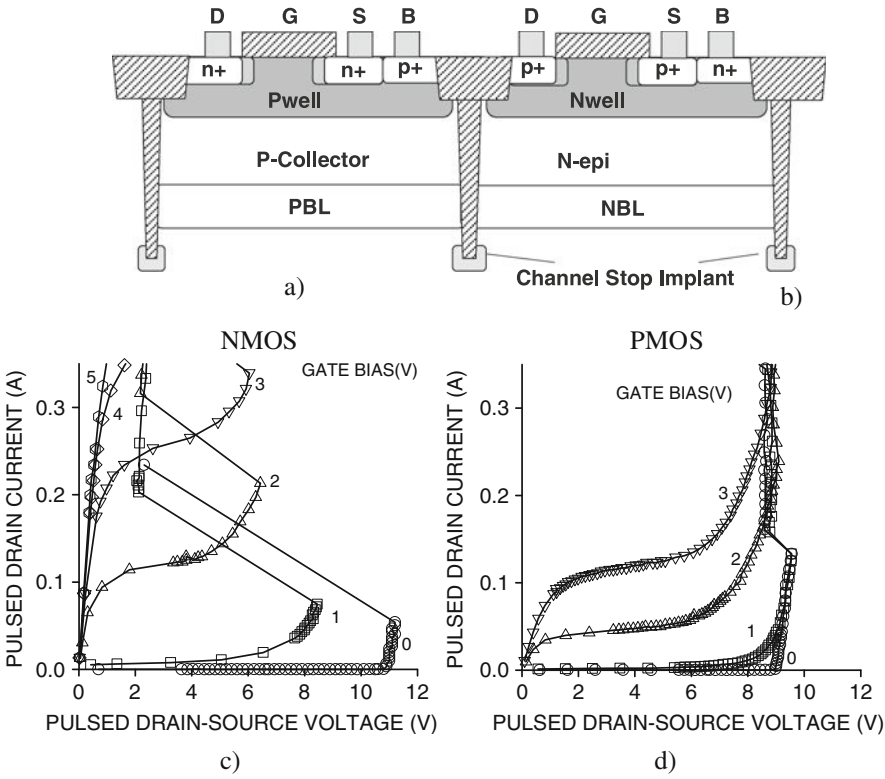


Fig. 3.13 Example of BCD process cross-sections and ESD SOA for LV NMOS (**a**, **c**), and PMOS (**b**, **d**), respectively

TLP characteristics of the device at different gate bias are presented in (Fig. 3.13c, d). In NMOS, the critical voltage for current instability is highly dependent upon the gate bias. Due to avalanche multiplication of the channel current, the instability condition $\alpha M>1$ is true at much lower multiplication coefficients M, which correspond to a lower drain voltage (Fig. 3.13c). In PMOS devices, the instability condition is much less dependent upon the channel current, due to the different conditions for avalanche–injection at the reversed mobility ratio for majority and minority carriers in the Nwell base (Fig. 3.13d).

In the case of NMOS device, the channel current determines the critical voltage dependence upon the gate bias and controls the turn-on of the parasitic n–p–n structure in the conductivity modulation mode. Since the ballasting region is absent, the device fails after snapback due to the possibility of unlimited current. The destruction results in the local contact and semiconductor melting due to the current filamentation effect.

Similar parasitic NPN and PNP structures can be recognized in high-voltage devices.

The channel region on the top of PBODY diffusion forms a high-voltage NLDMOS device (Fig. 2.14). Unlike the similar-in-purpose Pwell diffusion in NDeMOS devices, this PBODY diffusion is self-aligned with the left gate edge and thus provides stable characteristics for the MOS part of the device.

This second diffusion to the n+-source is essentially the reason for double-diffusion MOS (DMOS) in the name of the device. The lateral device is further combined with the drift region NDRIFT, which provides a similar drain extension function as the Nwell of the dedicated NEXT implant in the drain-extended MOS device. The region is critical in providing the combination of a high breakdown voltage and low hot carrier degradation parameters, since at the applied high voltage, the space charge region occupies most of NDRIFT. At the same time, the most important characteristics for a power device, on-state resistance, is directly related to drift region design. Different RESURF (reduced surface field) methods [46] are used to improve device characteristics, starting with field electrodes, field planes and junction geometry, to different multi-RESURF or super-junction [46, 47] solutions.

Thus, parameters of the gate length, NDRIFT-POLY overlap (X), and drift region length can control device SOA in a wide range.

Similar design considerations are true for the PLDMOS.

Unlike relatively short channel LV NMOS, NLDMOS has an embedded drift region that is similar to a saturation resistor. That is, after avalanche breakdown, the device may operate in avalanche mode with an additional voltage drop on the drift region.

Typically, in high-voltage standard devices, irreversible failure immediately follows the triggering into snapback [48] (Fig. 2.14). Thus, the boundary line that connects the critical voltage regime represents a current instability boundary. This instability boundary limits ESD SOA and can be, respectively, applied as a reference to devices of all sizes. However, the device topology may significantly change SOA.

The ESD SOA can be further used to establish the ESD protection window as a function of the control electrode coupling provided by the internal circuit during ESD stress. Thus, the voltage range between the maximum operation voltage of the integrated circuit pin and pulsed SOA is defined with consideration of the suitable ESD protection clamp that provides snapback in this range.

One of the hardware-level difficulties for TLP-pulsed SOA measurements is the functional test. In an ideal case, the conventional TLP leakage test can be automated to allow a functional leakage test between zaps at zero gate bias.

The cross-sections and examples of SOA for representative 20 V BJT devices are presented in Fig. 3.15 [49].

This particular example presents BJT with a polyemitter that is formed on the top of a shallow-base implant.

Measurement of the pulsed SOA in BJT devices presents methodological difficulties, since the voltage or current from the dc power source cannot be supported in the pulsed regime. Thus, except in shorted and open base cases, it is difficult to be sure about which conditions on the base correspond to a particular TLP plot.

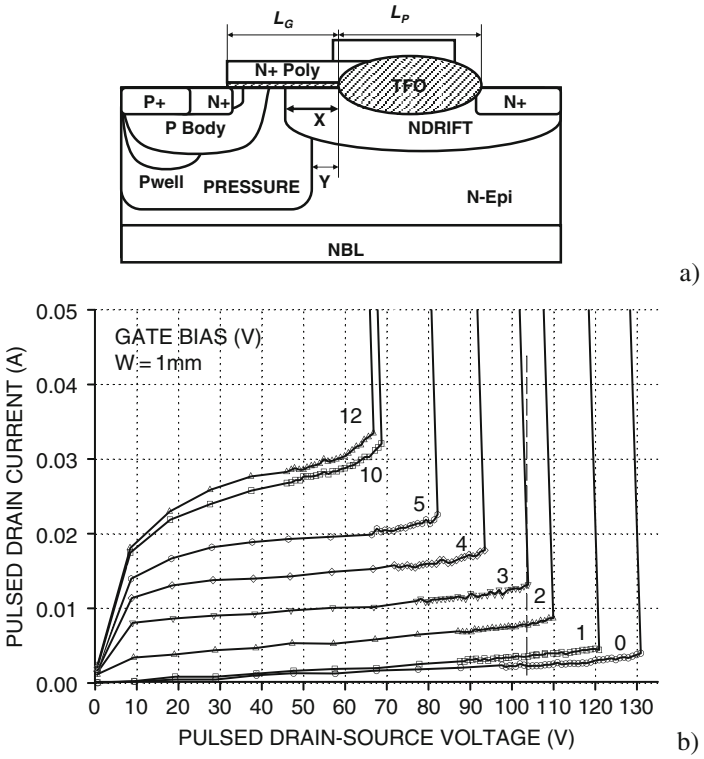


Fig. 3.14 Cross-sections for 100 V NLDMOS (**a**) and measured pulsed SOA for 100 V NLDMOS (**b**)

Nevertheless, the instability boundary is obtained according to experimental data, and thus can be used to estimate the ESD protection window range (Figs. 3.13, 3.14 and 3.15).

3.2.3.1 Waveform SOA Measurements

An alternative method of evaluating SOA is based upon the HBM waveform capture technique [33]. The tools for this technique have recently become commercially available [32]. A single pulse waveform tester allows measurement of real ESD pulse current and voltage waveforms.

This type of SOA measurement presents more complete information. As can be seen, in addition to the information similar to TLP in the 70–90 ns domain, the phase *I–V* characteristic represents both the overshoot and turn-off parts.

The main outcome from pulsed SOA characterization is accounting for the effect of the control electrode coupling. Indeed, under general ESD test conditions, the internal circuit may supply unknown conditions on the control electrodes of devices connected to the analog circuit pin. For example, in the case of the NMOS device

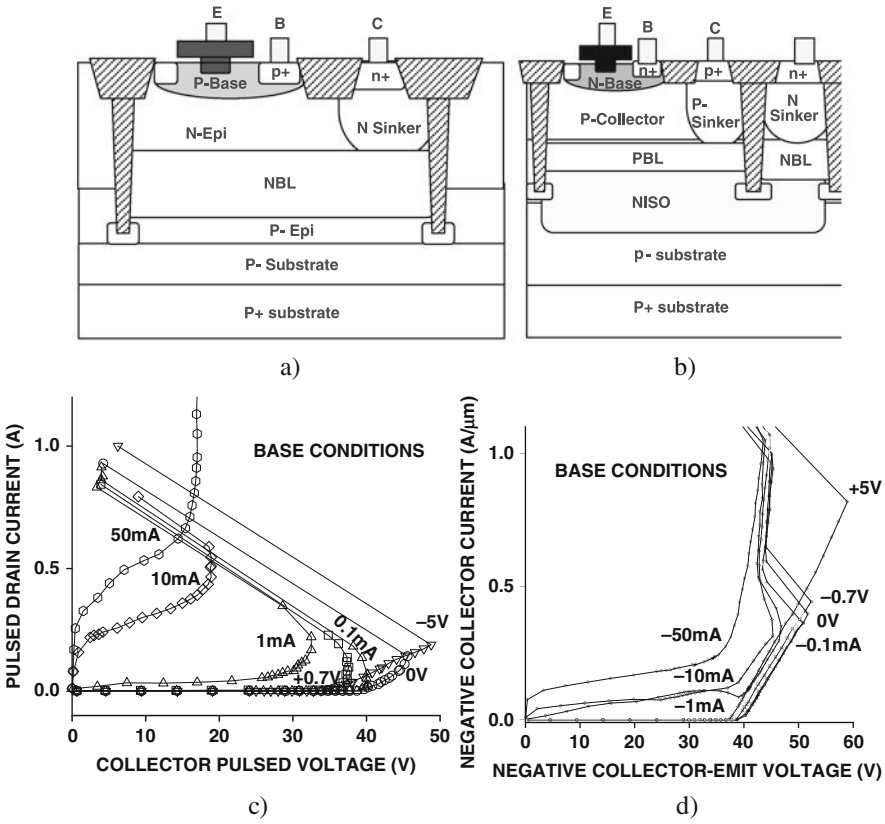


Fig. 3.15 Simplified cross-sections (**a**, **b**) and experimental pulsed SOA for typical 20 V NPN (**c**), 20 V PNP (**d**), respectively

(Fig. 3.13a) at zero gate bias, the maximum voltage is much higher in comparison with the on state conditions.

3.2.4 Instability Boundary and SOA for ESD devices

In the case of standard devices, the ESD SOA is mainly required to ensure a sufficient voltage range for local ESD protection implementation by choosing an ESD clamp with the appropriate triggering voltage. In a different way, the SOA data can be used to calculate an appropriate separation resistor for second-stage ESD protection, which allows use of the reversible current provided by the internal device for internal voltage clamping (Chapter 4).

Excluding self-protection cases (Section 3.5), standard devices usually fail after entering into snapback. This phenomenon is observed due to the absence of current-limiting regions in most device structures. In some cases, if lateral cur-

rent transport at the surface is disabled, vertical NPN devices provide reversible snapback operation due to internal subcollector ballasting.

Thus, in general, pulsed ESD SOA for standard devices supported in integrated process technology is coincident with the current instability boundary. When the electrical regimes reach the boundary, the positive feedback in the device structure initiates a certain type of conductivity modulation, followed by the current filamentation effect and creation of conditions sufficient for local heating and burnout.

This statement is not true for ESD devices. The ESD device is intentionally designed to support reversible pulsed snapback operation in the conductivity modulation mode. Therefore, in properly designed devices, the instability boundary does not correspond to irreversible failure.

The way to achieve such operation is a direct subject of the ESD device (Sections 3.3, 3.4, and 3.5), topology, and clamp circuit design (Chapter 4). However, the TLP or HBM waveform measurements are useful in establishing control of the ESD device. This information can further be used to design ESD clamps with a biasing circuit or to take advantage of d*V*/d*t* coupling of the control electrode. An example of pulsed characteristics for NLDMOS-SCR is presented in Fig. 3.16.

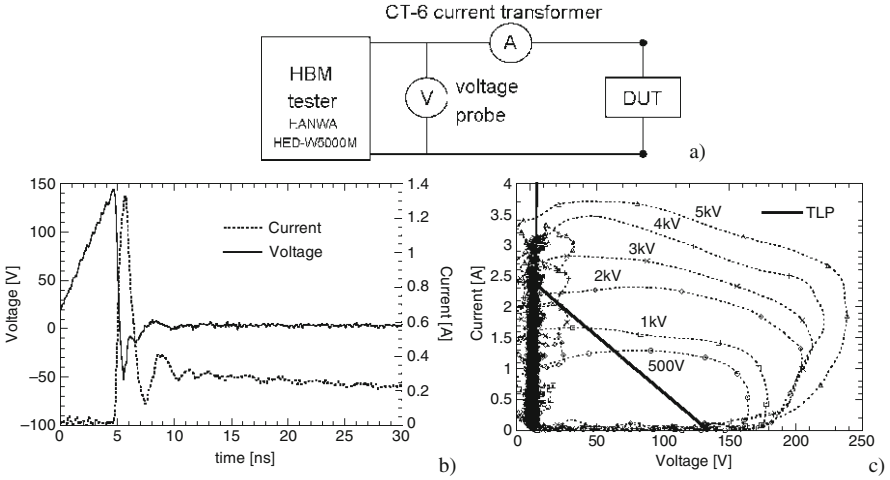


Fig. 3.16 Experimental set up for waveform SOA evaluation (**a**) and typical output for the waveform obtained for the 100 V NLDMOS device for the HBM pulse voltage waveform (**b**) and plotted phase *I*–*V* characteristics (**c**)

Similarly to the standard devices certain understanding of the ESD SOA for ESD devices is useful. After snapback in high-current state, device characteristics at different gate bias are practically merged. This fact reflects a rather simple conclusion that in high injection conductivity modulation regime, the device is practically uncontrollable, due to in-depth current conduction. As a result, in MOS devices, the effect of surface control electrode becomes negligible. A similar situation is realized in BJT devices, due to high injection conditions.

For snapback ESD devices, SOA is usually confined in the vicinity of one critical regime. This regime corresponds to the next stage of conductivity modulation. Thus, reversible operation of ESD devices in secondary breakdown mode presents a "normal" ESD operation regime.

Physical limitations of the operation regime for the ESD device itself are related to the same physical effects as those that affect standard devices [9]. However, these effects are usually realized at high current due to the conductivity modulation of the contact regions, rather than the active and drift regions of the device. In this case, the modulation is occurring between the contact metal and the semiconductor part of the structure in high injection state. This creates a physical equivalent of parasitic structures M-i-n+ or M-i-p+.

In case of semiconductor structure failure, the problem can be reduced to the conductivity modulation of the contact regions. This scenario of evolution results in much higher injection from the metal electrodes. The positive feedback results in the exceeding of contact and metallization limits and a localized heating and melting due to the current filamentation effect.

These phenomena are related to modulation of the contact region, fast electrothermal instabilities in the pulsed regime due to high-power heating of the active layers, and damage of the silicon structure by optical phonons and pieso-electric effects [9]. In opposite, the irreversible failure scenario involves electromigration limits of the contacts and metallization that may finally act similar to a fuse and create an open circuit event. A similar effect might happen in the case of bond wire failure. If backend limitations are removed, in principle, the physical limits of the ESD device operation are related to the energy dissipation in the silicon material itself, under the condition of typical heat generation within the few-microns deep subsurface regions. Some aspects of such limitations are discussed below.

3.2.5 Physical Limitation of ESD Devices. Spatial Thermal Runaway

A complete scenario of electro-thermal conductivity modulation that includes current and spatial electrical and thermal instability for the case of ESD operation has been demonstrated in [50]. The case of SiGe BJT devices in 0.24 μm BiCMOS processes has been studied.

According to the experimental TLP collector-emitter *I–V* characteristic (Fig. 3.17a), local burnout has been observed at some critical pulsed current (Fig. 3.17b).

In order to explain phenomena that limit device operation in high-conductivity mode in the case of ESD time domain, analysis has been completed using a thermal coupled mixed-mode simulation for the finite element device with geometry and implant profiles of a 0.24 μm BiCMOS process (Fig. 3.18a). A quasi 3D simulation approach (Section 2.7) has been applied to a distributed multiple-cell structure. The distributed structure has been "assembled" from segments of the original BJT cross-section obtained from calibrated physical process simulation

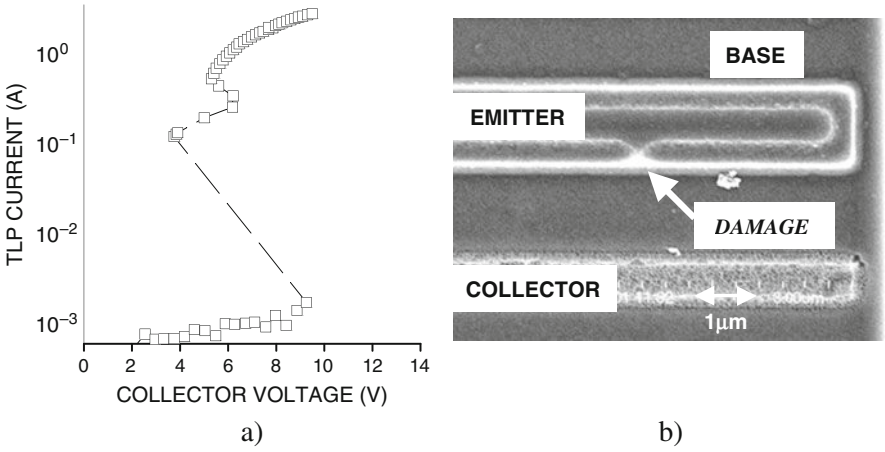


Fig. 3.17 TLP snapback curves for BJT pulse operation and SEM photos of the damaged device

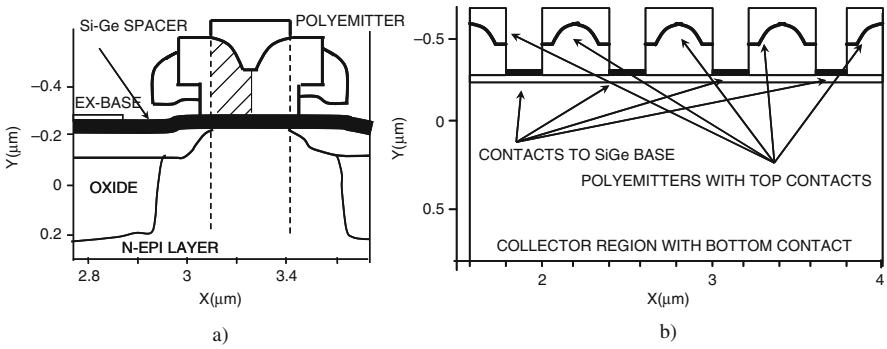


Fig. 3.18 Cross-section of the original NPN BJT (**a**) and part of 9.6 μm distributed structure (**b**) obtained by multiple reflections [50]

flow (Fig. 3.18b), thus providing a mathematical approximation for the general 3D problem and representing the case of multifinger BJT array.

For visualization of the current stratification effect in transient conditions, the maximum lattice temperature values have been extracted for the sequence of 1 μm regions across the Z=10 μm structure width. These dependencies of the values $T_1 \ldots T_{10}$ upon time have been compared for the cut lines with $Z_1 \ldots Z_{10}$=1...10 μm, respectively. At the uniform temperature distribution, all these dependencies were coincident. The appearance of the difference between them indicated formation of a spatially non-uniform solution.

The results of the thermal coupled mixed-mode simulation in the quasi-3D structure have been compared to the 2D cross-section of the BJT device (Fig. 3.18a).

In the case of 2D simulation for the Si–Ge NPN BJT cross-section, the device is triggered in the snapback regime after reaching a critical triggering voltage

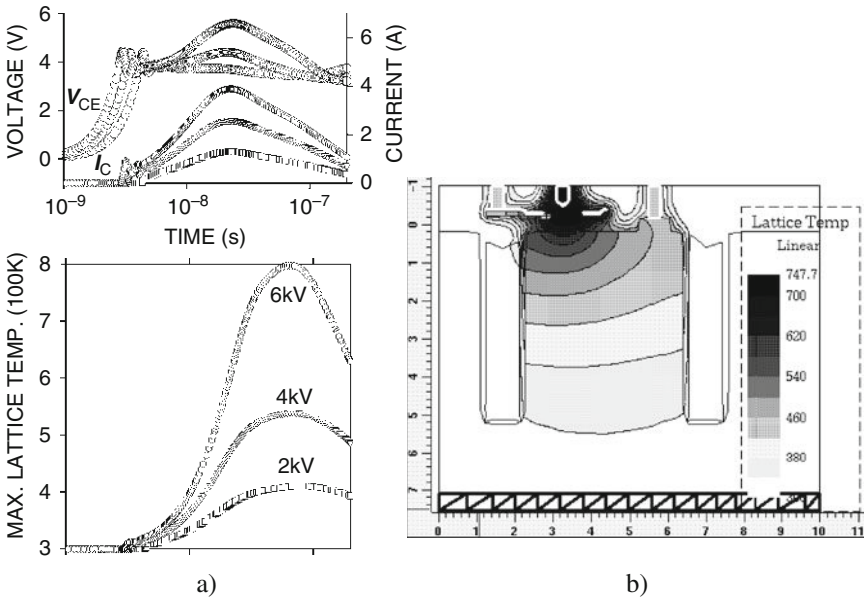


Fig. 3.19 Transient characteristics for the ESD protection clamp with BJT structure (Fig. 3.18) at area factor 200 for different HBM pulse amplitudes (**a**) and the lattice temperature distribution in the 2D device after 100 ns from the beginning of the 6 kV HBM pulse (**b**) [50]

(Fig. 3.19). Due to a very short ESD time domain, the heat generation is localized in a very narrow vicinity of the emitter-base region (~1 μm) (Fig. 3.19b). With increase of the ESD pulse amplitude, the peak temperature increases (Fig. 3.19a) according to the applied power to the device and at some point can reach the critical level for local structure melting or other type of material damage. However, in this 2D case, the current density along the imagined finger width is assumed uniform. Thus, the results of this simulation present a significant overestimation in comparison with the real effects that involve the short pulse electro-thermal current filamentation effect. This effect has been revealed using a solution for the quasi 3D structure.

In the quasi 3D structure, current filamentation is allowed and essentially proves that real physical limitation occurs at a much lower level. The quasi-3D BJT structure (Fig. 3.17b) generally provides a uniform response to ESD stress only in the range of HBM ESD pulse amplitude below a certain critical level. In this case, the solution is uniform across the sample width and all extracted temperatures are the same $T_1=T_2=T_3...=T_{11}$ (Fig. 3.20a). Above some critical ESD pulse amplitude, a different scenario is observed.

In this case, the extracted values of maximum lattice temperatures at different Z coordinates in the different 1 μm parts of the device indicate splitting of the uniform numerical solution (Fig. 3.20b), thus indicating a current stratification effect.

The current stratification effect results in the formation of a solitary hot spot (Fig. 3.21). This hot spot is formed at a much lower HBM pulse level than in the 2D

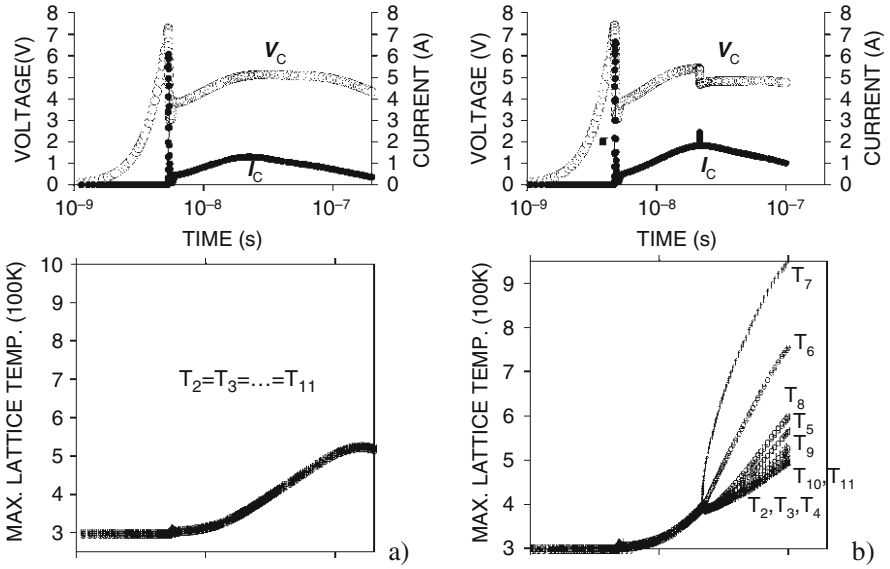


Fig. 3.20 Mixed-mode transient ESD characteristics for the protection clamp with the equivalent W=9.6 μm distributed BJT structure at area factor 42 for 2 kV (**a**) and 2.8 kV HBM pulse (**b**)

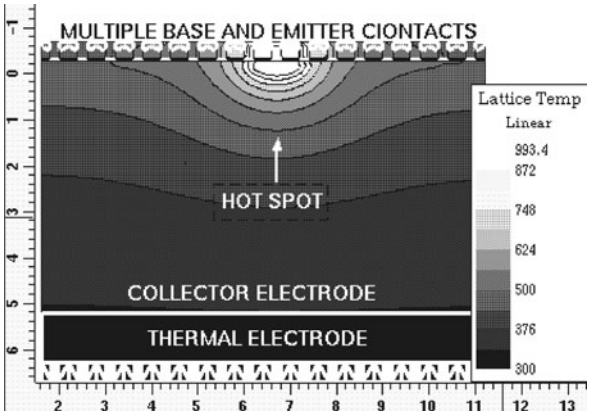


Fig. 3.21 The depth profiles for lattice temperature distribution in the distributed BJT equivalent structure after 100 ns of 2.8 kV HBM ESD stress

simulation structure that assumes uniform current distribution (Fig. 3.19). A critical regime for such electrothermal current filament formation is observed (∼HBM 2.7 kV), which corresponds to the uniform distribution of the maximum temperature of ∼ 700 K along the device finger (in the X-direction).

As soon as current filament is formed, a catastrophic event should be expected, due to enormous peak lattice temperature of >1500 K generated in the filament region (Fig. 3.21).

3.3 Low-Voltage ESD Devices in CMOS Processes

Under significant simplification, the ESD device design principle can be explained in a very straightforward way. It consists of implementation of a certain "parasitic" elementary structure that can support conductivity modulation (Chapter 2) inside either the supported device structure, or in a "free" ESD device, in the conditions of local current density limitation.

Referring to the previous SOA section, this practically means that the design of a new pulsed semiconductor device is expected to contain both the reversible instability boundary and desired instability boundary parameters.

The following sections of this chapter will discuss device-level design, while Chapter 4 will address the topological solutions referring to device layout.

The problem of proper ESD device design is mainly confined within several device engineering tasks

(i) provide an energy balance inside the device structure;
(ii) provide proper current balance at the clamp level that will enable width scaling;
(iii) obtain a solution that will combine triggering and holding voltage parameters and dc voltage tolerance, and produce an adequate voltage limitation waveform on the protected pin, while not interfering with the circuit during normal operation regime.

Practically pulsed ESD SOA for standard devices is required both to identify the spec for the desired ESD device and simultaneously to produce a starting point for free ESD device design.

One of the primary goals of device-level ESD development is to experimentally obtain a primitive ESD device to be used as a component in the future ESD clamp. Essentially, this means to obtain an ESD device by adding new regions and making transformations in the supported integrated device. This approach is demonstrated in this section for low-voltage devices.

At the same time, there are no direct guidelines that define low-voltage and high-voltage devices. For example, in the CMOS dual gate oxide process, low-voltage devices with a thin gate oxide are usually called low-voltage, while the devices with a thick gate oxide and corresponding high-voltage tolerance are called high voltage. This definition has been used for CMOS process description in Section 3.1.

The following section will use a different definition for high- and low-voltage devices. Since the majority of current analog products are based on 0.18–0.5 μm-process technology, we will rely on one of the most conventional definitions for the LV and HV in this book. A voltage of ~12 V will be taken as the dividing line between the low- and high-voltage ESD devices. This voltage level approximately corresponds to the n+-Pwell and p+-Nwell breakdown voltage in 0.13–0.5 μm processes.

3.3.1 Snapback NMOS

Historically, the snapback NMOS ESD device was widely used in integrated CMOS processes as a component of the grounded gate NMOS clamp (Fig. 3.22). This device is based upon the avalanche–injection conductivity modulation mechanism in the parasitic n–p–n structure that is already present in the NMOS device. This n–p–n structure is formed by the source acting as an emitter, the drain acting as a collector, and the Pwell acting as base with a corresponding p^+-tap contact region. The base length and gain of the n–p–n structure is controlled by the gate length L_G, while the internal base resistance R_{PW} is determined by bulk contact to source spacing L_{SWS} (Fig. 3.22).

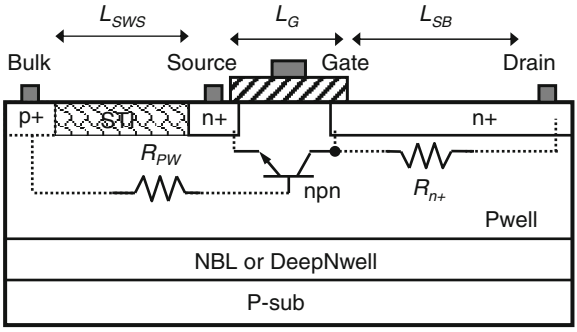


Fig. 3.22 Cross-section for snapback NMOS device in a non-silicided CMOS process with a parasitic n–p–n structure

In triple well CMOS process optional substrate isolation can be achieved by the DeepNwell region with corresponding lateral Nwell isolation. In case of BiCMOS process N-buried layer can be used for the same purpose.

In spite of simplicity, this device combines several important design features.

First of all, the gate electrode can be coupled to the source through an additional resistor R_{GS}. During the fast rise time of the ESD pulse, the gate electrode is biased above the threshold voltage by the current through the C_{GS}/C_{GD} capacitive divider that creates a voltage drop on the resistor, where C_{GS} and C_{GD} are the gate-source and the gate-drain parasitic capacitances. In accordance with pulsed SOA (Fig. 3.13a), due to the multiplication of the channel current, the conductivity modulation is realized at a lower multiplication coefficient, thus providing a lower triggering voltage in comparison with the dc breakdown voltage.

Thus, snapback NMOS in grounded gate configuration represents a transient triggering solution. In addition the bulk coupling effect provides for the dV/dt turn-on at a lower voltage level than expected from the supported similar NMOS device with the same drain diffusion length. Due to the long ballasting region, this effect is significantly stronger than in standard devices with symmetrical source and drain regions. A gate resistor is required to disable the clamp at the normal operation region. However, the gate coupling in this clamp should be considered at a high-speed I/O design to avoid parasitic current conduction.

The second feature is the drain ballasting region. The region provides saturation resistor R_{n+} with the parameters controlled by the ballasting region length L_{SB} (Fig. 3.22). This region is required to limit local drain current in the conductivity modulation mode, thus providing a reversible operation in a wide range of the ballasting region length (usually 1–8 μm). To obtain the region in silicided processes, a silicide exclusion SALEX mask is usually used. It is, however, important to provide appropriate mask alignment that will completely remove silicide at the poly edge, as discussed in Section 3.1.

Finally, the base (Pwell) diffusion region placement defines the internal base resistance of the parasitic n–p–n device. This parameter is critical in achieving avalanche–injection instability conditions.

Thus, snapback NMOS ESD device operation is based on the parasitic n–p–n structure, with additional current induced by the channel electron current.

NMOS clamp layout design specific will be discussed in Chapter 4.

3.3.1.1 Three-Dimensional Simulation of Current Instability in Snapback NMOS Devices

The physical limitation of the snapback NMOS device is similar to the pulsed electro-thermal effects discussed previously for the case of Si–Ge NPN BJT.

The original solution for physical effects that result in catastrophic failure of the NMOS structure has been obtained using a full 3D simulation [50].

The case of the rather simplified 3D NMOS structure with a 10 μm contact width (Fig. 3.23a, b) has been studied. Calculated ESD HBM transient characteristics of this simulation object are presented in Fig. 3.23c. Starting from HBM pulse amplitude of ~ 6 kV, the maximum lattice temperature becomes non-uniform, representing the current filament and corresponding hot spot formation phenomenon. A cut-line along the drain junction under the gate demonstrates the spatial electro-thermal current instability of the drain current evolving with time, as well as the solitary hot spot in the structure (Fig. 3.23). Kinetics of current stratification is presented in Fig. 3.24.

Both NPN BJT and NMOS devices have the same physical limitation of ESD device operation regime. It is related to a rather fast thermo-electrical current filamentation of the ballasting subcollector or drain regions. At the same time, before the critical level, the avalanche–injection conductivity modulation is balanced below the critical level by either the sub-collector or drain ballasting region.

Typically, the current limit of the ESD device operation itself is not a limiting factor of ESD protection capability for a given circuit pin. Depending on design, the ESD device can withstand rather high-current levels at high voltage. Therefore, before the damage of the ESD device itself, the internal circuit can fail due to excessive voltage at the pin. Thus, specifically the waveform of the ESD voltage limitation produced by the ESD device is a major concern. Roughly, the waveform is projected from the TLP measurement as a function of the triggering and holding voltages and on-state resistance of the device in the high-current state after snapback.

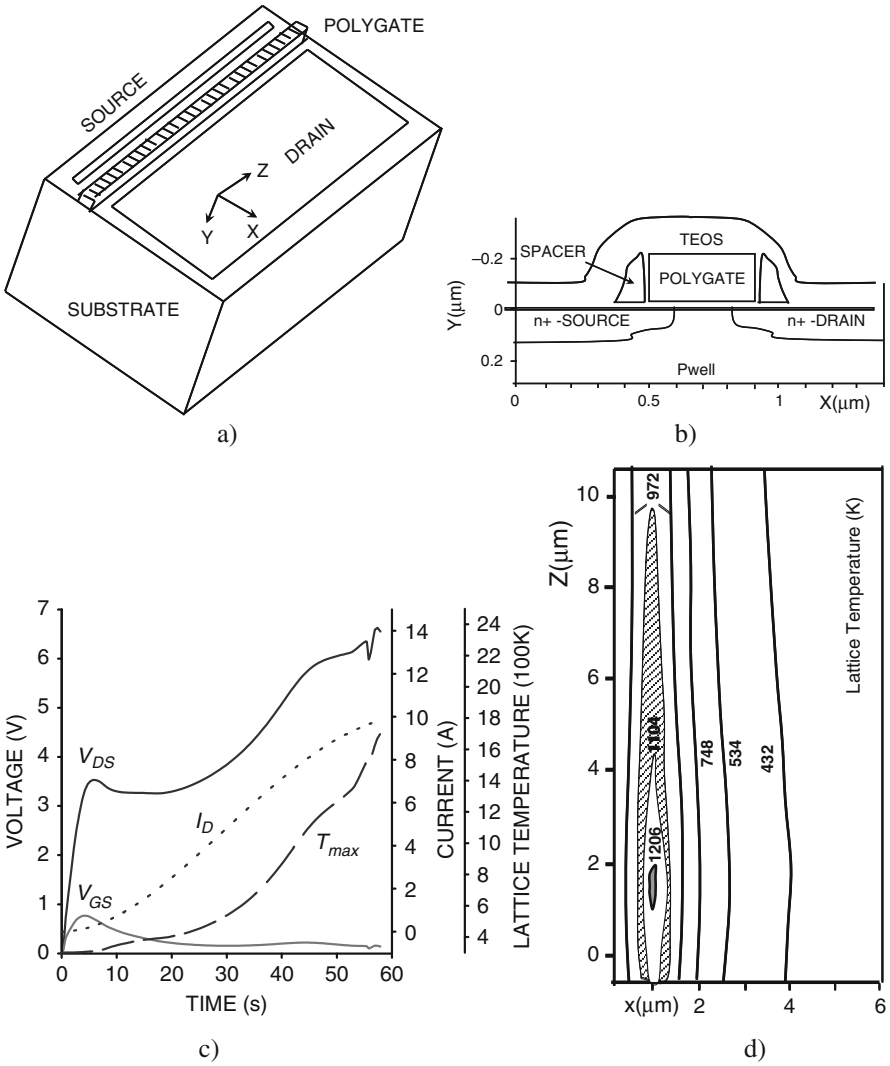


Fig. 3.23 Three-dimensional NMOS structure (**a**) and the cut plane in the gate region vicinity (**b**). Transient ESD characteristics (the drain voltage, the drain current, and the maximum lattice temperature) of the 10 μm 3D NMOS structure with an area factor of 50 for 10 kV HBM (**c**) and the cut plane representing hot spot formation at Y=2 μm (**d**)

3.3.2 FOX (TFO) ESD device

The voltage tolerance of the snapback NMOS is limited by the maximum rating of the gate oxide. Therefore, in case of higher voltage tolerance, a different device-level solution is required, depending on the voltage level. In some cases stacked

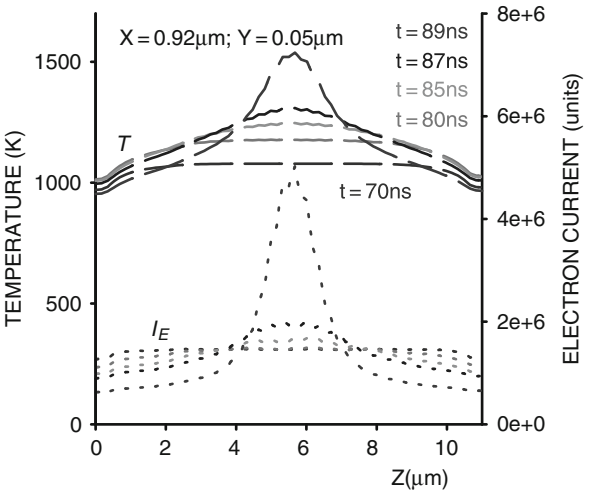


Fig. 3.24 Kinetics of current stratification obtained from the 3D simulation solution

clamps can be exploited. However, if I/O or power supply voltage is above the gate oxide limits, the best suitable structure is the so-called field-oxide (FOX), thick field oxide (TFO), or STI device. The device presents itself as a free lateral NPN, formed by CMOS diffusion and the surface isolation region in the Pwell.

A typical example of such device application is the EEPROM write pins. These pins usually require an elevated voltage level above the gate oxide maximum reliability rating, for example, 6–7 V in 5 V NMOS devices. While the snapback NMOS device is capable of providing an appropriate triggering voltage, it is no longer useful, due to reliability constrains at normal operation regime.

In the case of the CMOS process, FOX is the next most logical ESD device structure. The FOX device can be obtained through transformation of the snapback NMOS device, by swapping the polygate region with a minimum dimension surface isolation region, for example, in the shallow trench isolation with a length of L_{STI} ~0.25–05 μm (Fig. 3.25). Often, names similar to original snapback NMOS are

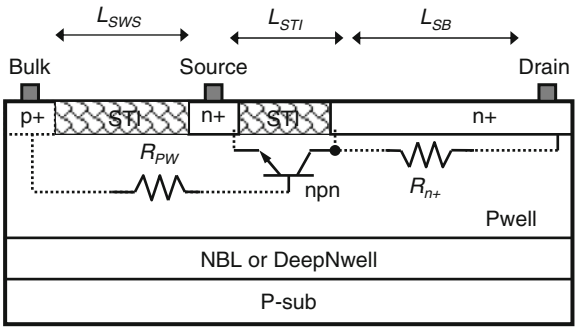


Fig. 3.25 Cross-section for snapback FOX device in CMOS process with indicated parasitic n–p–n structure

preserved for the electrodes and regions of this device. In this case, the n-emitter can still be called as source, the collector region as drain, while the P-base is called a P-well or a Bulk.

Similarly to snapback NMOS, the device is self-aligned, since the n+-implant is not penetrating through the surface isolation oxide. The remaining design features are the same as in the snapback NMOS. To limit the current density increase in the conductivity modulation regime, the device requires the n^+-ballasting region on the local level. This region is usually selected based upon experimental results in the same range of ~1–6 μm, depending on the Pwell and n+-doping regions. The well tap placement is designed with large spacing to the source (n-emitter) of the structure to provide a high intrinsic Pwell (p-base) resistance, thus controlling the triggering voltage.

In the case of SOI processes, the Pwell contact can be treated similarly to the base of NPN. In this case, taking the advantage of substrate isolation, the device can be connected in a common emitter circuit with an additional resistor and optional voltage reference device (see Chapter 4).

There are two ways to tune the device turn-on in high-conductivity state: (i) by using the dV/dt effect occurring due to displacement current in the p-base-n-collector (drain-well) junction in ESD regime; and (ii) by adjusting the implant profile at the collector region to control the breakdown voltage of the device.

The FOX device presents a more compact solution in comparison with the stacked device. Often, a precise adjustment of the breakdown and triggering voltage is required. One of the practical ways to implement such a solution is to use the Nwell implant to create a graded junction (Fig. 3.26a). The major concern in conjunction with this solution is non-self-aligned device architecture and understanding of the triggering voltage variation upon Nwell position. It is important to understand the impact of the Nwell–Pwell masks misalignment. An example of the experimental implementation of the HV FOX device with a graded Nwell (n+)-Pwell junction is presented in Fig. 3.26b. Within standard misalignment for a process of ~0.1 μm, the 12 V voltage tolerance can reliably provide a deviation within ~11–13 V.

Both the snapback NMOS and the FOX device have similar parasitic n–p–n structures with relatively low-current gain due to a lengthy p-base (p-well) of ~0.5 μm. Therefore, the results and understanding of Chapter 2 is directly applicable in describing the physical effects of avalanche–injection conductivity modulation and spatial current distribution in these devices. To reduce the threshold of instability, a large internal and often external base resistance is implemented. To limit the current density on the local level, a distributed n+-ballasting region is used.

3.3.2.1 Surface NPN

An alternative self-aligned device can be engineered as a surface NPN using a corresponding n+-mask (Fig. 3.26c). For high voltage, a self-aligned n+–n+ spacing can control the device triggering characteristics. For low voltage, if the process allows, a PLDD blanket implant can be done at the surface.

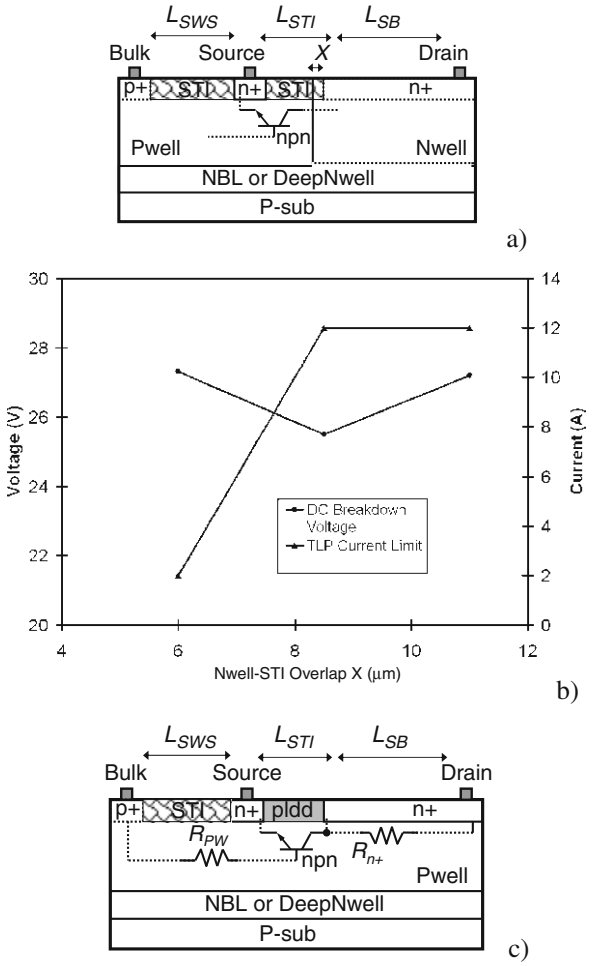


Fig. 3.26 Cross-section snapback FOX ESD device with graded Nwell junction (**a**) and dependence of the breakdown and triggering voltage as a function of Nwell overlap parameter *X* (**b**). Implementation of the surface NPN device with PLDD (**c**)

The practical problem of such devices is the unstable characteristics that are very dependent upon process variation, due to formation of the surface states. The surface states can impact the device breakdown voltage by walkout effects. The walkout effect can be realized as a result of charge trapping in the lightly doped surface region. Another limitation is the minimum space between two n+- implants, which might be too big for appropriate device parameters to be achieved.

A disadvantage of the device is the worse heat dissipation conditions, due to confinement of the current at the silicon surface.

3.3.3 LVTSCR and FOXSCR

The next class of snapback ESD devices compatible with CMOS processes is the silicon-controlled rectifier (SCR) devices. These devices combine a parasitic n–p–n and parasitic p–n–p structure that form an internal parasitic p–n–p–n structure (Chapter 2). Therefore, the numerical analysis and phenomenological description are applicable to the analysis of the physical processes in these devices.

There are several ways to obtain SCR in the CMOS process. In the NMOS starting structure, the device transformation requires implementation of the p+-emitter region isolated by the Nwell (Fig. 3.27a). In the case of the PMOS starting structure, the implementation requires additional substrate isolation by deep N-well or Niso regions. In this case, the Pwell-isolated n+-emitter can be embedded into the device in order to form the p–n–p–n parasitic structure with SCR properties (Fig. 3.27c). The Pwell that is isolating the n+-emitter requires not only vertical isolation from the p-substrate but also lateral isolation by corresponding Nwell regions.

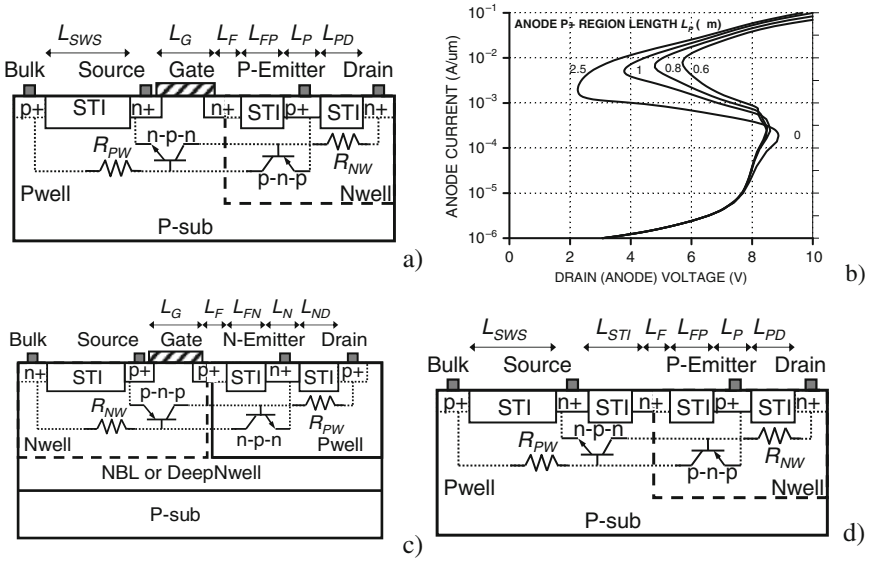


Fig. 3.27 Cross-section for LVTSCR (**a**) in a 0.5 μm process and simulated *I–V* characteristics at gate bias 1 V for different P-emitter length (**b**). The cross-section of p-LVTSCR device (**c**) and FOXSCR (**d**)

Historically, one of the most popular implementations of the SCR device for the CMOS process has been suggested in [51] with the name low-voltage silicon controlled rectifier (LVTSCR). This device can be obtained as a free device in typical CMOS process technology using a simple modification of the snapback NMOS within limitations of the process layout design rules. The LVTSCR device incorporates additional p–n–p structure achieved by the introduction of an isolated p+-emitter region. As a result of this modification, the new equivalent circuit of the

device includes both parasitic n–p–n and p–n–p structures with well resistors R_{NW} and R_{PW} (Fig. 3.27a). In addition to the L_{SWS} and L_G parameters that control the characteristics of the n–p–n structures, several new parameters control the characteristics of the p–n–p structure. These parameters include the floating drain region L_F, the p-emitter length L_P, and corresponding spacing L_{FP} and L_{PD} (Fig. 3.27a). An example of the holding voltage control in the LVTSCR device by the anode region length is presented in Fig. 3.27b. According to simulation results, both the triggering current and holding voltage can be substantially changed by the reduction of the gain of the p–n–p structure.

The p-emitter isolated by Nwell can approach the n+-source rather close and avoid punch through. At the same time, the NMOS structure provides the starting current in the avalanche multiplication region at the drain junction. Therefore, to combine both SCR and NMOS in a single device, a floating n+- region is required to separate the Nwell from the n+-source.

The n+-floating drain region solves the misalignment problem as well. Small variations of the Nwell position across the floating drain region within mask alignment tolerance provide only a minor variation of the breakdown and triggering characteristics of the device. These characteristics are defined by the NMOS structure.

The principle of LVTSCR operation consists in turning on the device into the double injection conductivity mode with a – typical for this mode – very low holding voltage of ~1.5–2 V (Fig. 3.27b). Thus, the device provides completely different voltage waveforms in comparison with snapback NMOS or FOX devices. Due to a much lower holding voltage, during the ESD pulse, the voltage is limited at a much lower level.

The process of triggering into double injection conductivity mode depends upon the device design and implant profiles. In properly designed devices, a domino-like effect is expected and can be demonstrated by device simulation. This domino effect can be explained as follows.

At first, the fast rise time of ESD pulse provides gate coupling above the threshold voltage, which results in the formation of the accumulation channel and source-floating drain current conduction through the Nwell region toward the n+-contact. If the Nwell doping is low, this current can already lower the potential barrier at the p+-emitter-Nwell junction and cause corresponding hole injection toward the source electrode. If the Nwell isolation of the p+-emitter is high, then the avalanche–injection in the parasitic n–p–n structure will occur, first providing a much larger current to open the p+-emitter junction. Normally, for 0.5 μm CMOS processes, the turn-on of the n–p–n structure is observed first, followed by p–n–p turn-on. When both parasitic structures are in on-state, the critical condition for positive feedback is defined by the (2.20) (Chapter 2).

This domino effect is typical for other SCR devices [7].

The final state of the device in the high injection mode is supported at very low multiplication coefficients. Under these conditions, the avalanche current component, generated in the high electric field, is completely replaced in the structure by the injection of holes from p+-emitter region and electron injection from the n+-emitter (source) region, generated in the low electric field.

The most valuable practical advantage of the device is the very high-current density and low clamping voltage level, in comparison with the snapback NMOS that correspond to the holding voltage V_H of the device. However, in the case of power voltage $V_{DD}>V_H$, the possibility of transient latch-up should be accounted for, in case of short term electrical overloads at the power supply pin.

Respectively, the FOX-SCR can be obtained using the same transformation rules, by swapping the MOS part of the device with surface oxide isolation region (Fig. 3.27d). Similarly to FOX, the FOX-SCR devices provide a high dc voltage tolerance by elimination of the maximum rating limits for gate oxide reliability. The breakdown voltage of the floating n+ drain to Pwell junction determines the turn-on of the device. The same approach can be used for graded junctions to create higher-voltage-tolerant devices, using Nwell grading of the blocking junction.

3.3.4 Low-Voltage Avalanche Diodes

Low-voltage avalanche diodes are rather useful components of ESD protection. If the devices are not supported in the given CMOS process as standard devices, they can be obtained freely and validated by the test data, taking into account process variation across the wafer and from lot to lot.

At an appropriate doping level, the diode can be constructed as a self-aligned surface device formed by n+-pldd, p+-nldd, or nldd-pldd implants (Fig. 3.28). The first two diodes are practically self-aligned, since lightly doped drain (LDD) diffusion can overlap with corresponding contact n+ or p+ diffusions.

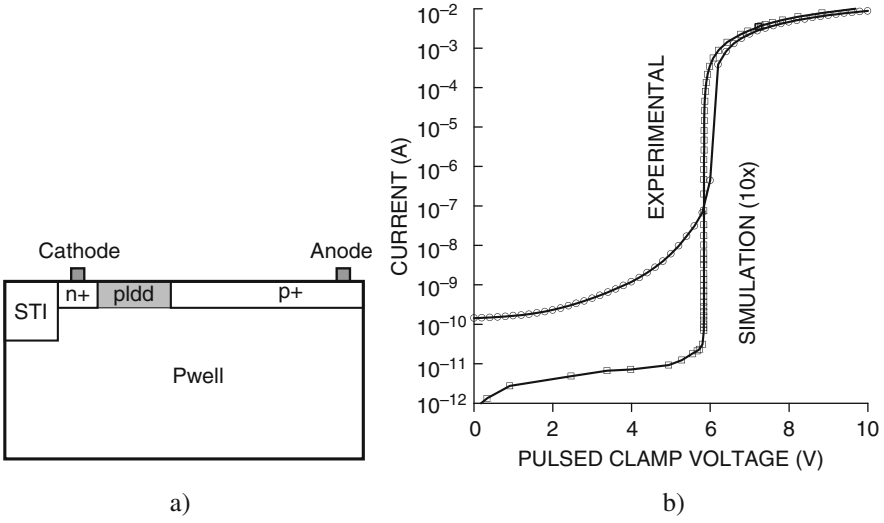


Fig. 3.28 Simplified cross-sections for n+-pldd surface avalanche diode obtained in CMOS process (**a**) and the experimental *I–V* characteristics (**b**)

A little technical difficulty in obtaining such a device is related to the CAD package automatic generation of LDD mask layers, and thus requires corresponding manual editing of this layer.

A much bigger obstacle is availability of an appropriate implant dose in the process that enables a truly low-leakage avalanche diode with a breakdown voltage of ~5–7 V. If the implant is too high, a tunneling component may significantly impact the diode breakdown characteristics, providing a lower breakdown voltage and higher leakage current at lower voltages. In an extreme case, a truly tunneling diode (or Zener diode [14]) might be formed, which can be inappropriate for the desired voltage range. The experimental *I–V* characteristics of the successful surface avalanche diode in 0.5 μm CMOS are presented in Fig. 3.28b.

As an alternative, in some cases the BiCMOS process may deliver an appropriate low-voltage avalanche diode that is formed by base and emitter regions of the BJT devices.

Building the non-self-aligned low-voltage avalanche diodes is usually impractical. Application of the low-voltage avalanche diodes in reference, stand-alone, or second-stage components of the ESD clamps is discussed in Chapter 4.

3.3.4.1 Comparison of the Surface and Buried Avalanche Diodes

Unlike CMOS processes, the BiCMOS and BCD processes may deliver an opportunity to build free low-voltage avalanche diodes using in-depth subcollector regions. Comparison of the surface and buried avalanche diodes is presented below.

Surface avalanche diode Type-I (Fig. 3.29a) is formed by the n+-region on top of the doping region DeepP. Another version of this diode, Type-II, is formed by the same regions under the n+-region and connected by the Pwell region (Fig. 3.29b). Due to having the same junction, both diodes deliver a similar breakdown voltage (Fig. 3.29c, d). However, the surface interaction effects provide higher leakage and pre-breakdown current. According to TCAD analysis, the reason for these effects is the breakdown close to the isolation region that results in involvement of surface states and traps, as well as the reduced surface field (RESURF) effect. In the Type-II avalanche diode, the breakdown is distributed more uniformly, due to the current flowing more in the vertical rather than the lateral direction.

Comparison of the buried and surface avalanche diodes is presented in Fig. 3.30b. The surface junctions provide a relatively low series resistance for the diodes, but the impact ionization process is originated at the surface. This can potentially make the breakdown voltage magnitude highly sensitive to the process-yield-related surface effects. Similarly, surface diode Type-II has the lowest series resistance value; the impact ionization process occurs at the edges of the N+ layer in the subsurface region. The magnitude of the breakdown voltage depends on the proximity of the breakdown region to the surface. A buried diode has an in-depth breakdown, but a higher series resistance.

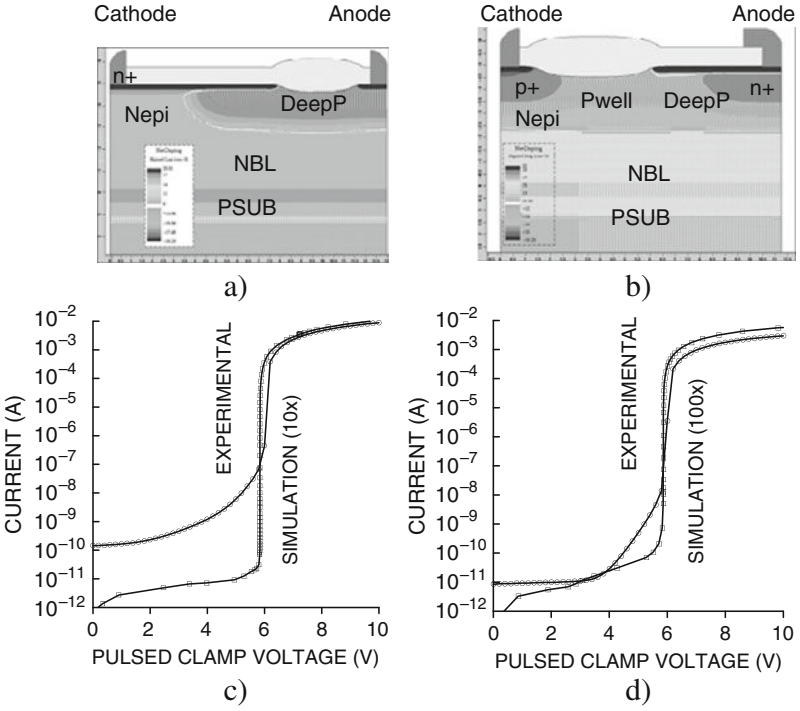


Fig. 3.29 TCAD cross-section (**a**, **b**) experimental and simulated *I–V* characteristics (**c**, **d**) for the Type-I and Type-II surface avalanche diodes

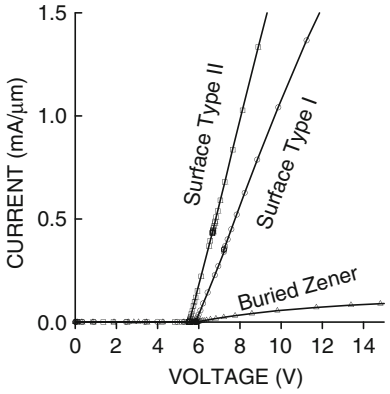


Fig. 3.30 Comparison of buried diode experimental TLP characteristics with two types of surface avalanche diodes

3.4 ESD Devices in BJT Processes

Apparently, ESD devices, discussed above, can be created in the CMOS part of the BiCMOS process too. This section discusses ESD devices specific for the BJT process.

3.4.1 Integrated NPN BJT Devices

Often, standard NPN BJT devices supported in the process can already provide reversible operation in snapback mode. In the opposite case, the reversible operation can be ensured by modification of the subcollector region parameters.

A typical NPN BJT cross-section in the 0.25 μm BiCMOS process is presented in Fig. 3.31. The device combines the base-emitter region with a thin intrinsic base and poly emitter, while the N-epi region and subcollector NBL and N-sinker regions form the collector region. Namely the subcollector regions with appropriate parameters, due to current saturation in these regions, are responsible for reversible operation of the BJT device in the conductivity modulation mode.

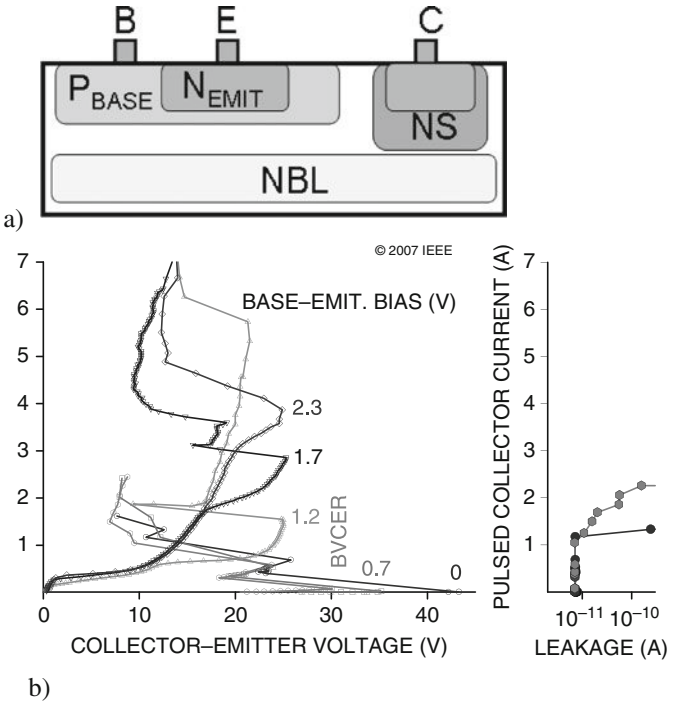


Fig. 3.31 Simplified cross-section for standard NPN BJT ESD device (**a**) and measured TLP collector–emitter characteristics at constant base-emitter bias (**b**) [96]

Under normal operation conditions, the collector current in the N-epi region is mainly vertical. In ESD conditions at high-current density, saturation in N-collector is observed, followed by lateral current transport.

In these conditions, a part of the NBL and usually a highly diffused N-sinker form an appropriate embedded ballasting region. The region locally limits current density in the NPN device during avalanche–injection positive feedback.

In principle, the physical processes during conductivity modulation in n–p–n structures which have already been discussed in Chapter 2 are directly applicable to the discussion.

The output collector–emitter characteristics for NPN BJT at different conditions in the base circuit are presented in Fig. 3.31b.

The turn-on voltage level for conductivity modulation mode of the first clamp is based upon two physical effects. The effects are represented by either the internal breakdown of the collector-base blocking junction or the dynamic effects. The dynamic dV/dt effects is related to the displacement current effect in the collector–base junction.

The collector–emitter 100 ns TLP characteristics have been measured according to the methodology described in Section 3.2.2.

ESD protection window boundaries for standard BJT can be estimated from the instability boundary (Fig. 3.31b), and corresponding parameters can be tuned for the ESD BJT device.

The BJT clamp-level solution is discussed in Chapter 4. Adjusting BJT triggering characteristics in a range that protects the standard device requires some alteration of the structure. Essentially, these measures are related to the implementation of the internal blocking junction, which has breakdown voltage lower than the collector–emitter breakdown of the standard device.

One of the ways to achieve this is to reduce the collector–emitter spacing, and thus create a different lateral junction profile using the tail of the N-sinker diffusion. The disadvantage of this solution is that the N-sinker implant is not self-aligned with base-emitter region, and significant misalignment can jeopardize the yield of the analog circuit.

The isothermal current instability that results in negative differential resistance in the collector–emitter I–V characteristics depends upon the product of the collector thickness L and the collector doping level N_D (Chapter 4). In the case of the collector–emitter breakdown with the open base (BVCEO regime) and the base current $I_B = 0$ the collector current is simply equal to the emitter current and is proportional to the value $(1 - \alpha M)^{-1}$.

When the emitter and base terminals are connected to the ground (BVCES) (U_{EB}=0, Fig. 3.31a), the initial location of the avalanche breakdown is at the collector–base junction. In this case, the avalanche breakdown is initially not accompanied by injection from the unbiased emitter junction. Therefore, the collector current is provided by the multiplication of the thermal generation current I_{C0}, i.e., $I_C \sim MI_{C0}$. In this case, the breakdown voltage is close to the collector–base breakdown voltage U_{CBBR}.

With the increasing avalanche current, the generated holes flow out of the base. In this negative base current regime, the hole current results in an additional voltage drop across the p-base, increasing the base potential at the emitter junction by the value $r_B I_C$, where r_B is the internal base resistance. The value of r_B can be extracted from the simulation or from measured data. Thus, when $r_B I_C$ is larger than or equal to the potential of the emitter junction opening (approximately 0.7 V), injection of electrons begins from the emitter and it is followed by the positive feedback mechanism for n–p–n structures described in Chapter 2.

3.4.2 Bipolar SCR

In addition to the standard BJT clamps, bipolar SCR [7, 52] clamps can be created. Simple device transformation rules require the formation of an additional p+-emitter region connected to the collector. Since by its architecture, NPN has an in-depth subcollector region, no additional regions are required to isolate the p-emitter (Fig. 2.32a).

To enable low critical current, the p+-emitter is introduced between the base region and collector contact region.

In principle, there is no architectural difference between the FOX-SCR and BSCR. The different naming is due to historical reasons. However, the major difference between the two consists in that BSCR is designed such that it is essentially composed of an original high-gain NPN device supported in process technology. In this case, the device repeats the characteristics of the standard device until current instability conditions are met.

The low-gain p–n–p structure embedded in BSCR by the introduced p+-emitter is formed by the N-collector acting as a base and the P-base region of the NPNP device acting as a collector.

The principles of ESD pulse operation for BSCR are similar to those of the FOX-SCR, as discussed above. The BSCR device provides a low clamping voltage that is typical for SCR, with triggering characteristics that are easily controlled by the base current (Fig. 2.32b).

Due to high internal gain, the BSCR device can produce a rather efficient dV/dt clamp that is triggered by the displacement current during rise time of the ESD pulse.

For both NPN and BSCR, an alternative version of the self-aligned ESD clamp with an internal avalanche diode can be obtained by stretching the PBASE diffusion to create an overlap with the collector region. However, this solution requires a significant deviation from the poly emitter NPN device architecture. Therefore, it is not possible in self-aligned BJT processes and Si–Ge processes with an epitaxial base.

A surface junction between the base and the N-sinker regions can be used to change the breakdown voltage of the device to adjust an ESD protection window targeting. Similarly, the internal avalanche diode structure can be implemented in the

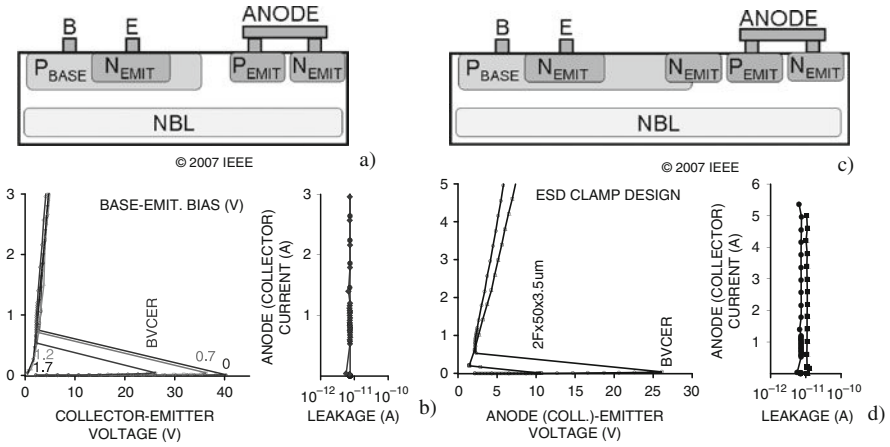


Fig. 3.32 Simplified cross-section (**a**) and TLP characteristics (**b**) of the original BSCR device at different base-emitter bias conditions and BSCRZ device with internal base-emitter avalanche junction (**c**) and its TLP characteristics compared to BSCR BVCER (**d**) [96]

BSCR, where an additional floating N-EMIT region is added between the PBASE and the P-EMITTER (Fig. 3.32c, d).

BSCR devices are usually designed as free devices based upon high-gain NPN, the current gain $\alpha_{NPN} >> \alpha_{PNP}$. The current instability criteria can be obtained, taking into account the additional gain from the built-in low-gain PNP α_{PNP} device. In this case, the current instability criteria with the grounded base is presented by $\alpha_{NPN} M_N + \alpha_{PNP} M_P > 1$. For the case of the BSCR, this term can be simplified to $\alpha_{NPN} M_N > 1$ due to a very low value of α_{PNP}.

Thus, in a properly designed BCSR, the instability boundary can be controlled similarly to the NPN BJT, making the design predictable through known yield characteristics for supported devices.

Depending on the design parameters, BSCR snapback is induced by avalanche–injection conductivity modulation, followed by the change of the modulation mechanism to double injection at high injection levels, which eliminates avalanche multiplication M_N, $M_P \sim 1$ and thus provides for the condition $\alpha_{NPN} + \alpha_{PNP} > 1$.

3.5 High-Voltage ESD Devices in BCD and Extended Voltage CMOS Processes

ESD protection of fast high-voltage pins in analog applications is a complex problem, due to the need to target several performance specifications. These specifications generally include appropriate voltage tolerance for the protected pins, precise turn-on voltage control below the pulsed absolute maximum voltage of the protected components during an ESD event, high-current operation for desired ESD pulse specifications, and low sensitivity to the pulse rise time.

It is also critical to maintain protection against ESD events during system-level operation. In this context, the most reliable way to guarantee system stability against transient latch-up is to require a minimum on-state holding voltage above the power supply voltage level. This requirement also maintains system stability against short-term electrical overstress (EOS) events above absolute maximum voltage limits.

Finally, since obtaining a small footprint device is a cost-driven requirement, it is obligatory for operation in the conductivity modulation (snapback) mode to achieve high current density.

High-voltage tolerant ESD protection may have two major specification requirements. In the case of high-voltage control pins or slow power pins, snapback devices with a holding voltage below power supply level can be used. However, in the case of high-voltage fast transient pins, hot plug-in or other system-level requirements, the expectation from a high-voltage ESD device is to provide high-current characteristics with the holding voltage above the power supply level.

In practical application, high-voltage ESD protection can be acomplished using high-voltage NPN devices based upon avalanche–injection conductivity modulation, SCR-type devices that provide high current with low holding voltage, and lateral PNP devices that operate in avalanche–injection mode.

Attempts to attain reversible snapback operation in high-voltage NDeMOS, NLDMOS, or VDMOS devices supported in the process are usually unsuccessful due to the irreversible nature of the snapback operation in such integrated components. The major reason for this is inappropriate current ballasting above the relatively low saturation current density of the high-voltage drift regions. Attempts to add an additional n+ ballasting region are also unsuccessful, due to deep current transport in the relatively long channel devices. In this case, current limiting is inefficient, since the current in the contact region is drawn from a much larger area, compared to the distance between the drain and source contact regions.

Similarly, active clamps with monopolar current conduction are inefficient due to the relatively high on-state resistance of high-voltage MOS devices. Therefore, snapback solutions are the solutions of choice for local protection of the high-voltage pins of analog circuits.

3.5.1 LDMOS-SCR and DeMOS-SCR Devices

As mentioned above, the most typical small footprint clamps are based on ESD protection devices operating in snapback mode. Either NPN BJT or silicon-controlled rectifier (SCR) ESD structures are generally used as high-current clamp components. Due to the nature of avalanche–injection conductivity modulation, both types of structures provide low holding voltage when implemented in high-voltage processes.

A typical cross-section of the DeMOS-SCR device implemented in a 0.5 μm extended voltage process is presented in Fig. 3.33a with the experimental TLP characteristics (Fig. 3.33b).

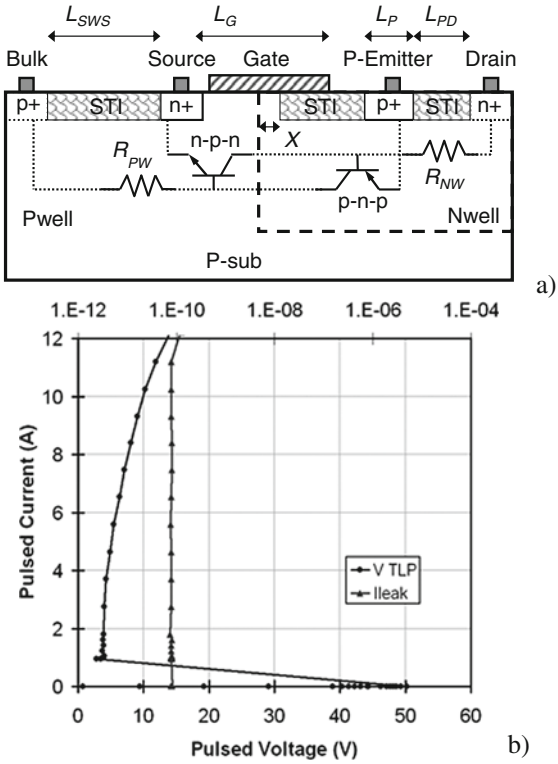


Fig. 3.33 Simplified cross-section (**a**) and the experimental TLP *I–V* characteristics (**b**) of typical DeMOS-SCR for 40 V ESD protection. [53]

The device combines a standard device supported by the process with parasitic n–p–n structure and an introduced low-gain parasitic p–n–p structure. The embedded p–n–p structure has a p+-emitter region with Nwell-region acting as a base and a Pwell p+ contact acting as a collector.

The principle of device operation is similar to the previously discussed LVTSCR device, with a corresponding change to the high-voltage blocking junction. Before snapback, the avalanche breakdown corresponds to the same operation mode as in the original LDMOS device. This is guaranteed, since the left part of the structure practically replicates the original device. After achieving avalanche–injection conditions, electron injection is initiated from the source junction in the parasitic n–p–n structure, with the n-collector formed by the Nwell drift region. At high voltage, the drift region is already almost fully depleted, according to the principle of LDMOS operation. Therefore, the drift current along the extended drain region creates an opening potential for the p+-emitter–Ndrift junction. As a result, at some critical current density, conditions for double injection conductivity modulation are created.

Similarly to LVTSCR, in order to adjust device turn-on, either dynamic coupling of the gate electrode or reduced drain-source (n-base to p-collector) spacing can be used in the real ESD clamp, which is discussed in the following chapter.

Control of the triggering and breakdown voltages can be achieved by changing the drift region length by the RESURF approach, thus shortening the space charge region area, for example, by using a drift region-poly overlap with a poly gate extension.

The floating drain region can be added depending on the depletion of the extended drain (drift) region the device. This can be done similar to NLDMOS-SCR (Fig. 3.34). The region might be required to avoid premature opening of the P-emitter junction, due to a small voltage drop across the drift region. Whether this is necessary or not depends on the doping profiles.

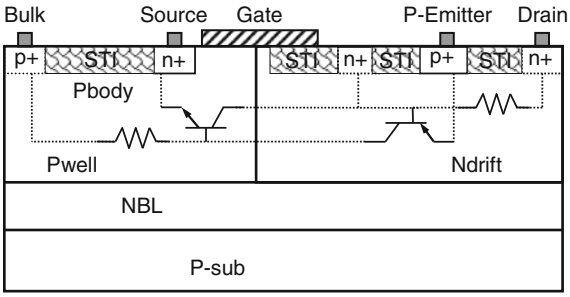


Fig. 3.34 Simplified cross-sections for high-voltage NLDMOS-SCR

Triggering current characteristics of the device depend on the p+-emitter isolation level. This can be controlled by the parameters of the floating n+- drain region that is soft-connected to the drain contact region either by the NDRIFT (Nwell), or by the n+-diffusion implemented by means of appropriate topology.

In the case of the BCD process, the device structure is practically similar to above with the exception of the N-epi and NBL (Fig. 3.34), as well as the Ndrift and Pbody, which play the same role as Nwell and Pwell in the NDeMOS-SCR. In principle, the device provides similar ESD characteristics and snapback operation with low holding voltage ~2–5 V. However, specific ESD clamp design is different in correlation with the differences in DeMOS and LDMOS layout (Chapter 4).

Applying similar transformation rules as in the case of a FOX-SCR device, a high-voltage SCR can be obtained by swapping the polygate region with the available type of surface oxide isolation region (Fig. 3.35). The device eliminates the MOS structure, but the poly region over the field oxide can still be used to create a field electrode.

Versions of PLDMOS-SCR implementation using PLDMOS as a starting structure are presented in Fig. 3.36. In this structure, an n+-emitter is added to create the parasitic components of the SCR device, while the blocking junction is supplied by the PLDMOS device architecture.

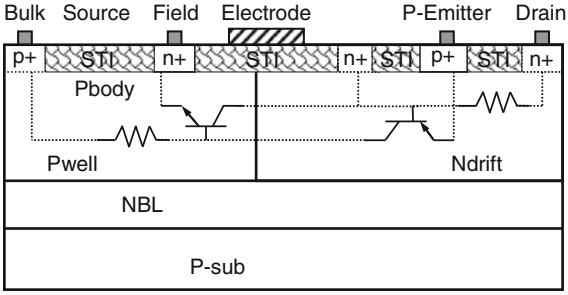


Fig. 3.35 Simplified cross-sections for high-voltage SCR

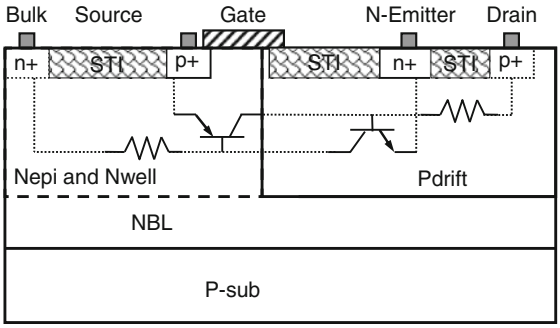


Fig. 3.36 Simplified cross-sections for PLDMOS-SCR devices

The common feature of all the high-voltage SCR devices (Figs. 3.33, 3.34, 3.35, and 3.36) is the embedded p–n–p and n–p–n internal structures which enable double injection conductivity modulation. At the same time, the dc voltage tolerance of the device is supported by the original standard device design that provides both the blocking junction and the induced channel current in case of gate electrode coupling or external gate control circuit. The topological control of the spacing of the device regions provides the control of the breakdown voltage, the triggering voltage, the triggering current, and even the holding voltage parameters of the composed ESD device.

3.5.2 Lateral PNP BJT Devices

If the space-on-chip budget is not critical, high-voltage protection can be achieved using large-size stacked clamps or by a high-impedance high-voltage avalanche diode.

However, if a small footprint is required by the design, perhaps the only universal practical alternative to low holding voltage SCR solutions is the lateral PNP device. The first successful PNP ESD device implementation has been reported in [54–56].

In these studies, it has been demonstrated that a PNP-based protection structure can be greatly improved to provide robust ESD protection for high-voltage I/Os in advanced smart power technology. By using specific design guidelines and coupling an open-base lateral PNP with a vertical avalanche diode, a low on-resistance (∼1 Ω) can be obtained using a limited silicon area. This optimized structure was used to protect specific high-voltage pins of a power-over-ethernet chip.

Later, implementation of the lateral PNP ESD device both in drain-extended voltage CMOS and in BCD processes with a supported lateral DMOS device has been demonstrated [53].

In [57], the physical mechanism that underlies the experimentally observed conductivity modulation in the lateral PNP has been studied using physical process and device simulation.

The structures and transmission line pulse (TLP) characteristics of the lateral PNP (Fig. 3.38a) can be compared to the DeMOS-SCR device (Fig. 3.33b) implemented in the same 0.5 μm 40 V drain extended CMOS process.

Architecture of the 40 V PNP device (Fig. 3.37) is based on the DeMOS device and can be obtained by transformation of the DeMOS-SCR, which involves elimination of the former n+-region and replacement of the polygate region and the n+-region with field oxide or shallow trench isolation. The extended drain diffusion region and Pwell form the blocking base–collector junction. The field electrode over the blocking junction (FE) (Fig. 3.37) is optional, but may be very important for the elimination of the undesirable breakdown voltage walkout effect [40].

This transformation removes the original parasitic NPN structure from the device, since the n+-source acting as an n+- emitter is removed. Therefore, conductivity modulation is now based upon avalanche–injection in the p–n–p structure. This structure provides a different current balance and the resulting high holding voltage (see Chapter 2).

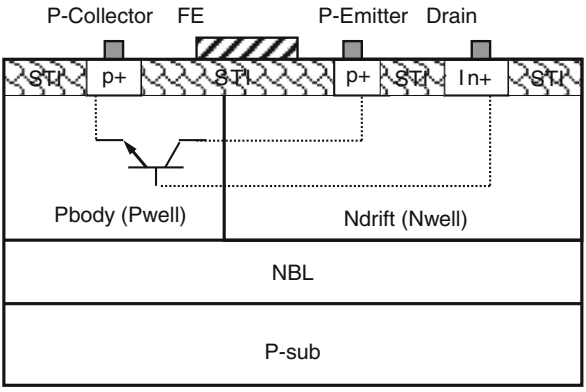


Fig. 3.37 Simplified cross-section of lateral PNP

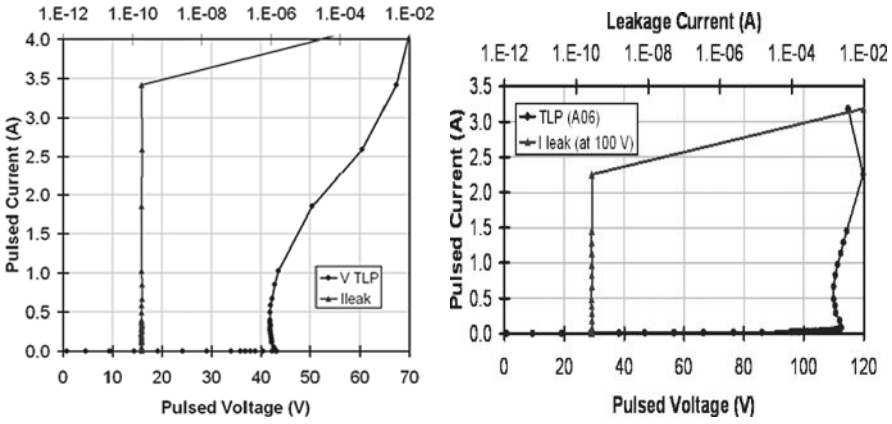


Fig. 3.38 The experimental TLP *I–V* characteristics for lateral PNP (device implemented in 40 V DeMOS process (**a**) [53] and for 100 V BCD process (**b**)

In comparison to the DeMOS-SCR, this PNP structure presents a much closer approximation of the desired clamping characteristics with a high holding voltage (Fig. 3.38a). The advantage of this device is the absence of the deep S-shaped dependence with the 3–5 V holding voltage of the DeMOS-SCR (Fig. 3.33b).

At the same time, due to the change in the majority carrier, the on-state resistance of the device is relatively high, and thus some negative differential resistance can provide for a better approximation of the desired ideal case.

The focus of [53] is on physical explanation of the experimentally observed phenomena for NDR in PNP, using Technology CAD analysis with calibrated physical process and device simulation tools. It has been demonstrated that specifically modulation of the extended N-drift (or Nwell) region is responsible for negative differential resistance of the lateral PNP devices, rather than the modulation of the Pwell region. This fact is also confirmed by the absence of the NDR in vertical PNP with a thin N-base region and corresponding long P-collector region.

Standard PNP devices show significantly different behavior in breakdown and conductivity modulation modes relative to their lateral counterparts [53]. Simulated 3.39a, b for the example of a 0.5 μm 20 V BiCMOS process. The *I–V* characteristics (Fig. 3.39c) of these devices are compared to simulated collector–emitter characteristics of a standard vertical NPN with an N-buried layer and an N-sinker subcollector.

The avalanche–injection conductivity modulation of the extended N-base region can be understood via comparative analysis of the 2D profiles and cut-line dependencies of the electric field magnitude at points α (switching voltage) and β (holding voltage) (Fig. 3.39). Electric field reduction is seen in the base (negative differential conductance) at holding voltage β in Fig. 3.39c. In the state at α, the electron density greatly exceeds the hole density in the base, while at β, the base is filled by a

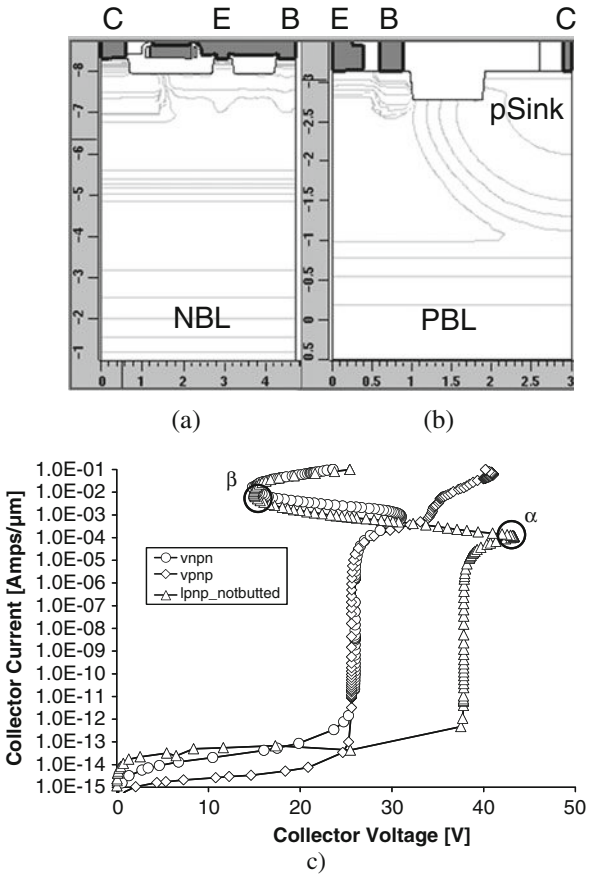


Fig. 3.39 Process simulated cross-sections for lateral (**a**) and standard vertical (**b**) PNP devices and their simulated collector–emitter *I–V* characteristics, in comparison with a standard vertical NPN BJT device in a common emitter circuit at zero base-emitter bias

quasi-neutral electron–hole plasma and the concentrations of the two types of carriers are approximately equal (Fig. 3.40d). This space charge compensation results in reduction of the electric field in the base (Fig. 3.40a) in the state at β.

The difference between electron and hole mobilities leads to significant difference in the quasi-neutral area formation of NPN and PNP devices, and the resulting decrease in the collector-emitter voltage.

A simple numerical simulation experiment can be used as a straightforward illustration of the differences in NPN and PNP behavior. The cross-section of a standard PNP is compared to the cross-section of an NPN, which is artificially created by swapping the n and p-doping of the PNP. Thus, an "antipode" NPN device is obtained from the PNP "ancestor." Another physical numerical experiment involves swapping all of the mobility coefficients between the electron and hole carriers in the

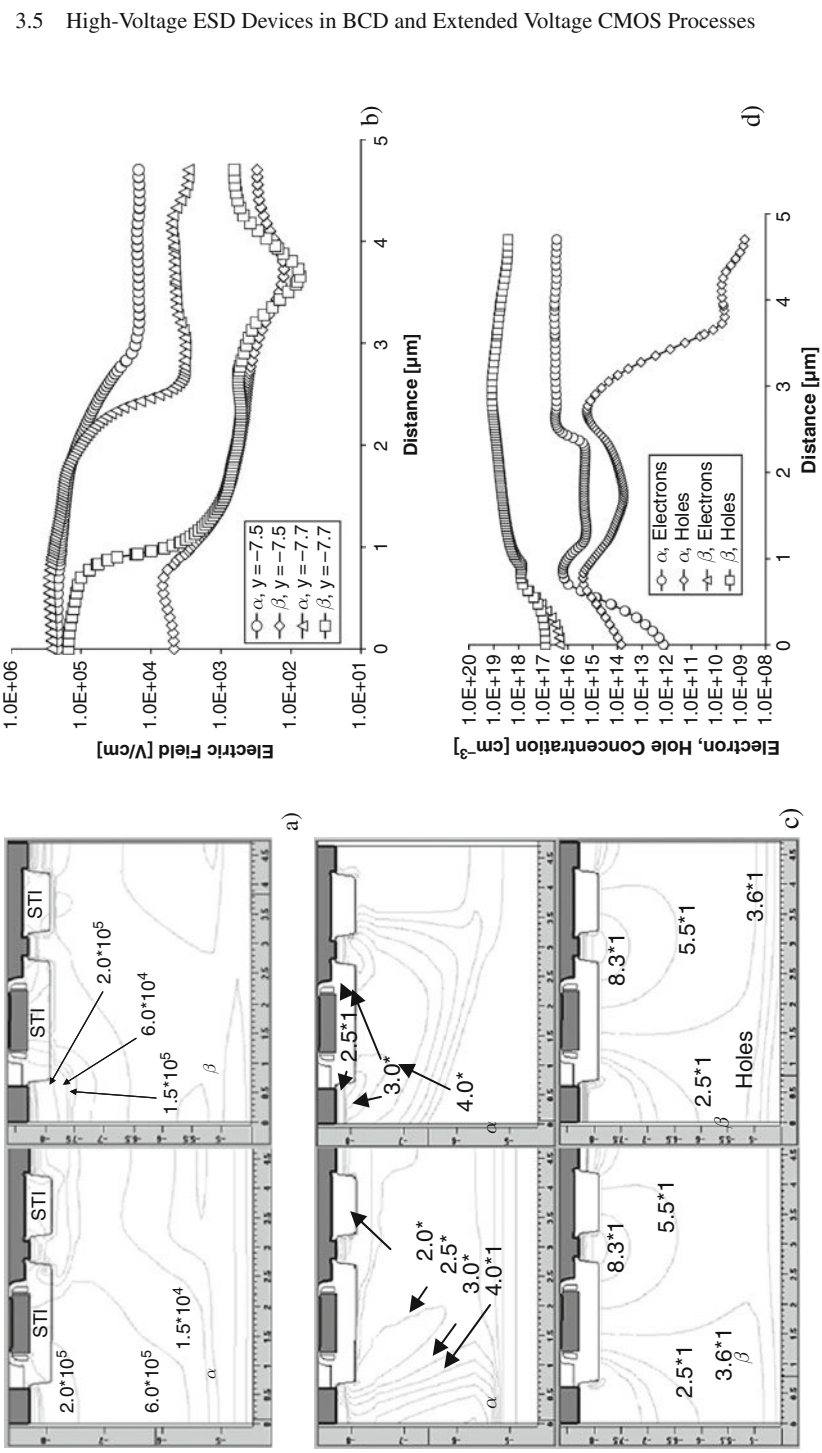


Fig. 3.40 Comparison of the 2D electric field magnitude profiles (**a**) with the corresponding dependencies at cut-lines (**b**) (at y= −7.7) and electron and hole density (**c**) with cut-lines (**d**) (at y= −7.7) for the states α (switching voltage) and β (holding voltage) indicated on the collector–emitter I–V characteristic in Fig. 3.39c

simulation. Both the device with the altered species and the device with the swapped mobilities demonstrate the same phenomenon – quasi-neutral region formation that provides a low holding voltage.

In [53], a mathematical model of the conductivity modulation in the high-voltage lateral PNP was proposed based on an equivalent circuit (Fig. 3.41), where an additional parasitic NPN is included in the PNP base–emitter circuit. This NPN is formed by the Ndrift region acting as a collector, P+-emitter as a base, and N+-base as an emitter, including corresponding shared regions.

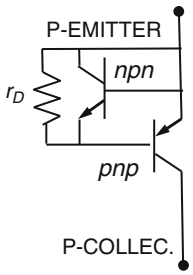


Fig. 3.41 Equivalent circuit of the high-voltage lateral PNP with the parasitic NPN device formed by PNP regions acting as Ndrift as collector, P+-emitter as P-base, and N+-base as N-emitter

3.5.3 High-Voltage Avalanche Diodes

High-voltage avalanche diodes are important reference components for dV/dt effect-independent high-voltage clamps.

Due to the fast avalanche process as compared to ESD pulse domain, the avalanche current can be used to provide bias or current for the control electrodes of the devices. Respectively, ESD devices can be designed with rather low coupling of the control electrode and injection junctions. The first can be achieved using resistors of smaller value in the common emitter circuit. The lower displacement current effect can be achieved by greater isolation of the injection regions. The greater isolation might involve use of a heavily doped isolation region and appropriate region dimensions.

As a result of such design, the high-voltage clamps with avalanche diode reference can provide dV/dt-independent turn-on, combined with high-current operation.

Similar to low-voltage design, depending on the options delivered by the process, either lateral or vertical high-voltage avalanche diodes can be implemented. In the case of the vertical diode, adjustment of the avalanche breakdown voltage is practically absent and the design mainly relies on what is already available in the process.

The lateral design presents a much larger degree of freedom in control of the breakdown voltage, by a simple mask change of the lateral dimensions of the diode. The disadvantage of lateral design is the non-self-aligned architecture.

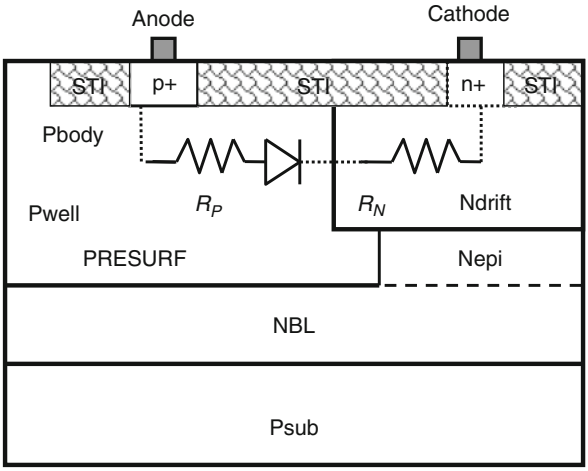


Fig. 3.42 Cross-sections for lateral high-voltage avalanche diodes in BCD process

A typical implementation of the avalanche diodes in BCD process reuses the regions used in NLDMOS devices (Fig. 3.42). The lateral high-voltage blocking junction is formed by the Pbody implant and Ndrift region. Since the Pbody implant is aligned to the STI position, the structure's sensitivity to misalignment is relatively low. The breakdown voltage of the diode is controlled by the Ndrift region length and optional PRESURF implant. The last is deeper than the N-drift region and creates a filed-plate effect, thus controlling the breakdown voltage of the device. The avalanche current generated by the device is limited by the saturation resistances of the Ndrift and Pbody regions R_N and R_P.

3.6 Dual Direction Devices

A small footprint and dual-direction functionality is required for a number of analog applications, for example, column drivers in liquid crystal displays, RF inputs, common mode voltage regulators, interface applications, and digital–analog converters. Design of dual-direction ESD protection is one of the most sophisticated device-level ESD problems, especially if system-level ESD protection requirements are also specified.

A typical requirement states that the protected dual-direction pin should be tolerant to a certain positive and negative voltage range relative to the grounded epi p+-substrate. Respectively, the ESD device is connected between the dual-direction pad and the ground node and must provide a proper waveform under the ESD pulse discharge.

Consequently, the device must include a corresponding double set of contact diffusion regions to realize a conductivity modulation mechanism in both directions, while at the same time, the regions should be separated by two blocking junctions.

3.6.1 Dual-Direction Device Architecture in CMOS Process

An absolutely necessary requirement for realizing a dual-direction ESD protection device in the CMOS process is the availability of a Deep Nwell or similar deep N-isolation layer in the process, which will vertically isolate the Pwell region. The exception to this rule is, of course, an SOI process case.

In [49], test structures have been fabricated using a non-silicided 0.5 μm 5 V CMOS process with a deep N-well, on a p+-substrate with a ~4 μm thick P-epi layer.

There are a number of options for arrangement of the contact diffusion and at least two ways to engineer the blocking junction: symmetrically (Fig. 3.43a) and asymmetrically (Fig. 3.43b).

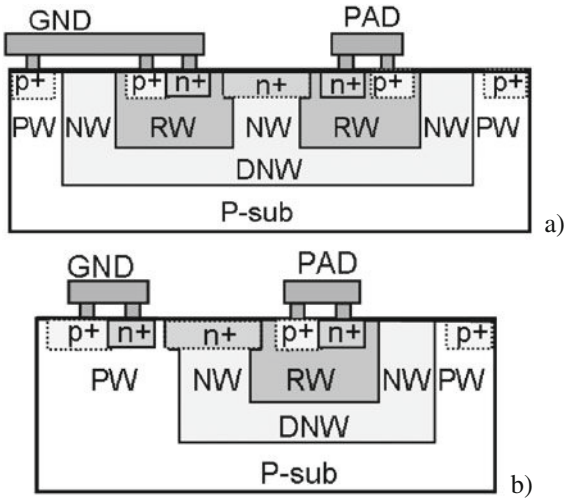


Fig. 3.43 Typical dual-direction SCR structure; PN-n-NP structure with the n+-Pwell (**a**), the compact asymmetrical version of the PN-n-PN device (**b**)

In the symmetrical case, the Nwell (NW) and deep Nwell regions (DNW) (Fig. 3.43a) can fully laterally and vertically isolate all Pwell regions (RW). This design minimizes the substrate effect described below.

In the asymmetrical case, only the pad side Pwell remains fully isolated (Fig. 3.43b). This solution is more compact; however, the substrate effect might significantly impact the operation of the device. The injection contact diffusions are connected to GND and PAD device electrodes.

Another feature generally implemented to reduce the voltage tolerance in the device is the middle n+ diffusion (Fig. 3.43). Since for a typical 5 V CMOS process, the P-well to N-well breakdown voltage is usually rather high (~18–25 V), the additional floating diffusion region is required to bring the blocking junction breakdown and corresponding triggering voltage closer to the 5 V core devices or the 10 V stacked device, i.e., to the ~10–14 V voltage range.

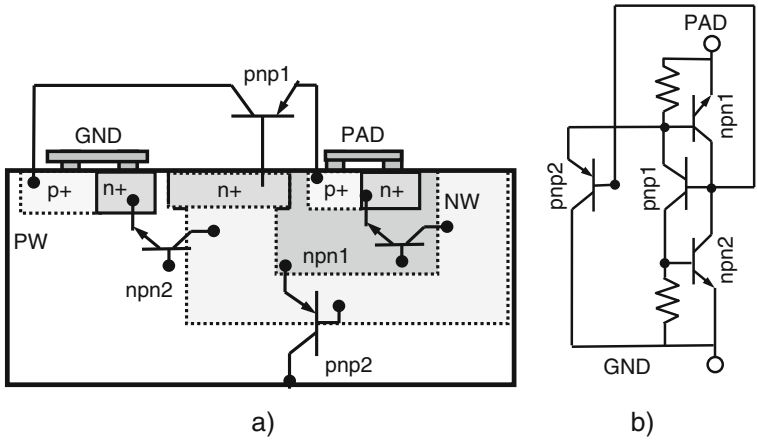


Fig. 3.44 Typical compact dual-direction SCR structure with the n+- Pwell and p+-Nwell blocking junctions (asymmetric PN-n-PN structure) (**a**) and its equivalent circuit (**b**) [36]

The device combines embedded parasitic n–p–n and p–n–p structures that can be identified in the device cross-section and represented by the equivalent circuit. This is presented in Fig. 3.44, for the case of asymmetrical device structure.

However, alternatively for each current direction, the injection regions form corresponding SCR-type internal p–n–p–n structures that enable double injection conductivity modulation in the device.

In [36], it has been shown that manipulation of the sequence of the injection regions or creation of a more complicated topology produced an important benefit of overcoming the undesired grounded substrate effect.

The problem consists in the embedded connection to the p-substrate by the guard ring necessary to provide latch-up isolation for the dual-direction ESD cell. As a result, the substrate acts as an additional p+- emitter in case of negative current through the pad, while in case of positive current, additional negative feedback is achieved.

3.6.1.1 Device-Level Positive and Negative Feedback

Thus, even symmetrical device (Fig. 3.43a) characteristics are significantly different, depending on the conditions of the floating or grounded p+- guard ring. In the case of a floating substrate connection, the device characteristics are symmetrical and provide a low holding voltage typical of SCR structures, while in the case of a grounded substrate, a significant asymmetry occurs (Fig.3.45). In the case of a grounded substrate, the device provides lower resistance in the negative direction, while in the positive direction, the holding voltage is rather high and the maximum current is reduced according to leakage measurement data.

This effect has been explained in [36] through the positive and negative feedback realized in the structure of the device. The asymmetry in the device operation can be

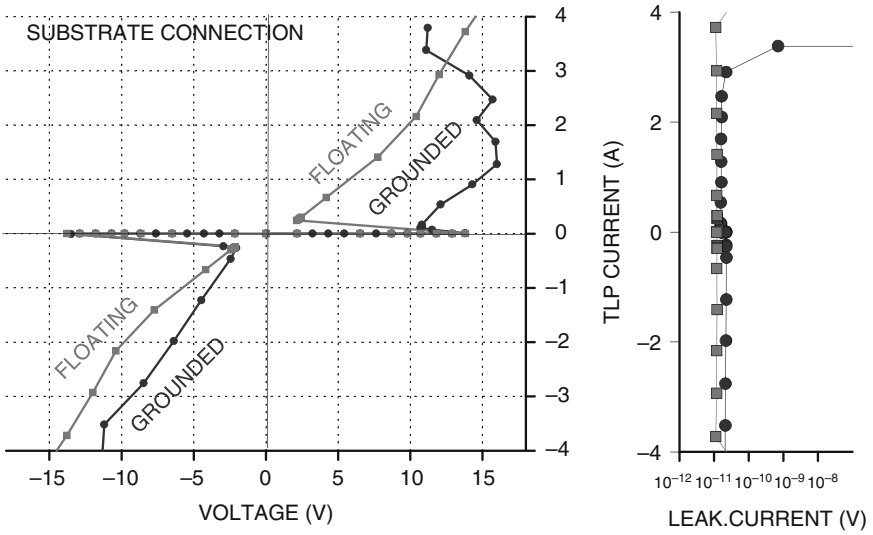


Fig. 3.45 TLP *I–V* characteristics of symmetric PN-n-NP structures for different substrate connection conditions

understood on the phenomenological level as a result of the engaging of a parasitic PNP. This PNP is formed by the p-substrate acting as a collector, deep N-well as a base, and the upper Pwell acting as an emitter.

After the device turns on into the conductivity modulation mode, in the case of positive current direction, the holes injected from the pad junction partially escape through the parasitic PNP current path into the P+-substrate. As a result of this effect, the level of the space charge neutralization between the electrons injected from the ground junction is reduced. This negative feedback results in the formation of a corresponding electric field distribution between the cathode and the anode of the DD-SCR structure and leads to rather high values of the holding voltage and the on-state resistance.

An opposite situation takes place in the case of the negative current direction. In this case, the p-substrate may essentially contribute to the hole injection level by acting as an additional P-emitter.

A similar effect can be observed with increase of the size of the symmetrical device. The effective collector resistance of the PNP increases as well as the negative feedback level is reduced, thus providing lower holding voltages for larger devices.

In order to compensate the device-level negative feedback provided by the parasitic PNP in the substrate, the diffusion sequence has been altered toward the PN-n-PN structure (Fig. 3.46a), followed by a reduction in the contact-to-contact spacing. With the positive current direction, the injection junctions are closer to each other.

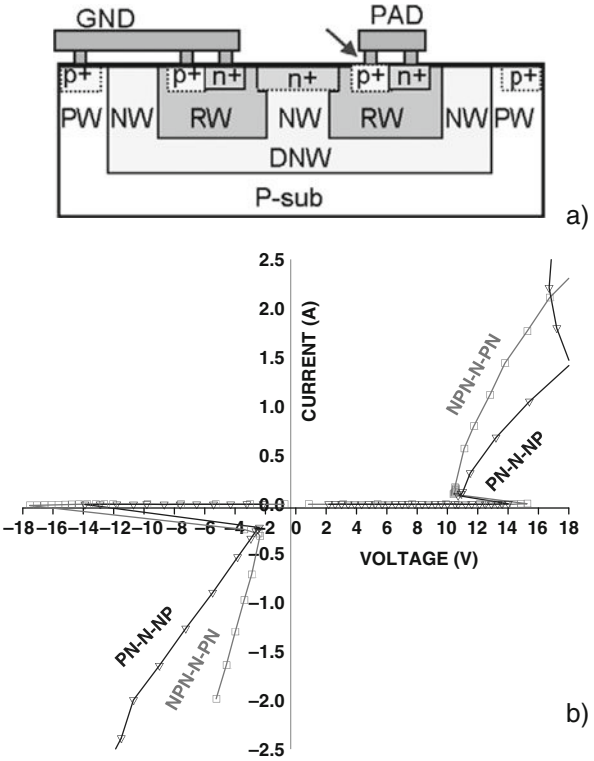


Fig. 3.46 Results of positive feedback implementation (**a**); TLP *I–V* characteristics of PN-n-NP structure (**b**)

This additional positive feedback mechanism compensates the negative feedback due to the parasitic PNP. At the same time, this measure does not impact the characteristics of the device in the negative current (Fig. 3.46b). For the case of NP-n-PN DIAC, both asymmetrical and symmetrical architecture types of the device perform similarly (Fig. 3.46b).

Reduction of the contact-to-contact spacing results in a further adjustment of the holding voltage within the desired range (Fig. 3.47).

Thus, it becomes very important to utilize knowledge about physical processes under conductivity modulation in ESD design, toward implementation of desired high-current characteristics of the device and addressing both the desired voltage waveforms and the potential dynamic latch-up issues.

3.6.2 High-Voltage Dual-Direction Devices

High-voltage versions of dual-direction devices are extremely sensitive to the positive and negative feedback effects described above.

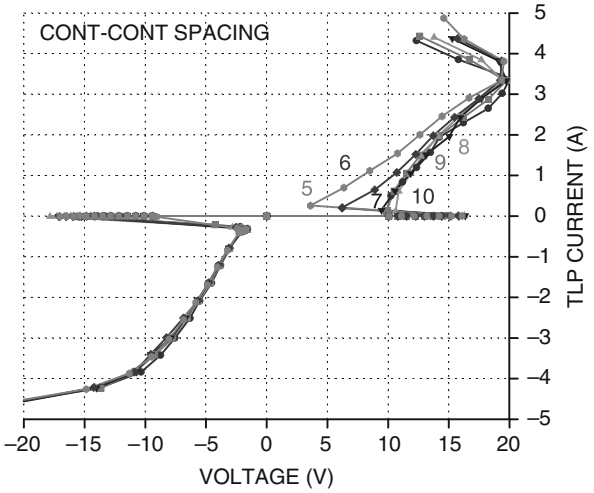


Fig. 3.47 Results of positive feedback implementation; TLP *I–V* characteristics of PN-n-NP structure created by scaling the device characteristics and the reduction of the contact-to-contact space

Usually, the challenging task is to implement both the high breakdown voltage and the high holding voltage when using these devices.

Substantial progress in this case can be achieved by a topological solution that is based upon formation of interdigitated contact regions. This approach is similar to the two-stage SCR for hot plug-in application described above.

The general architecture presents itself as a dual-directional SCR (DIAC type) that uses regions typical for BiCMOS process. The blocking junction of the device is formed by a Pwell-Nwell-Pwell sequence with a floating Nwell, or a PCOLL-Nwell-PCOLL sequence (Fig. 3.48).

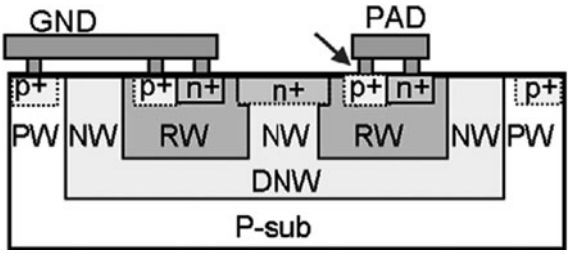


Fig. 3.48 Cross-section for the DIAC

There are several design variations of the interdigitated device: (n)P–N–P(n), where small n-regions are distributed mainly along the p+-diffusion, and (p)N–P–N(p), where small n-regions distributed mainly along the n+-diffusion (Fig. 3.49).

For the given device spacing, the holding voltage is the function of the balance of the diffusion regions for (n)P–N–P(n). An increase of the holding voltage from

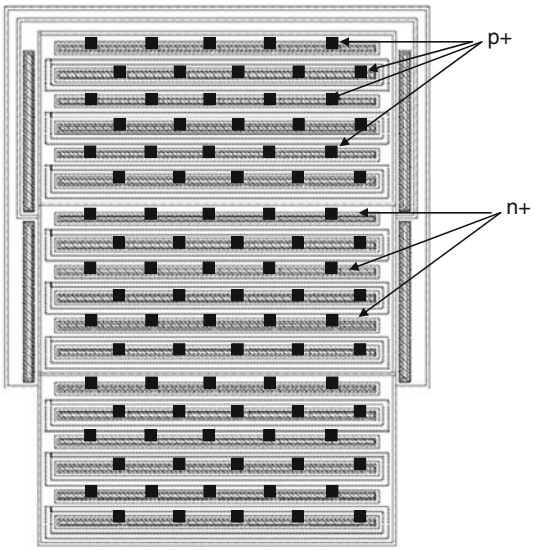


Fig. 3.49 Topology of the DIAC structures with interdigitated diffusion

the original holding voltage of ~5–7 V, up to 60–80 V (Fig. 3.50) can be achieved by using interdigitated regions with different area ratio.

However, a major side effect of the increase of the holding voltage is the reduction of the maximum holding current (I_{T2}) from I_{T2}=10–12 A (for V_H= 7 V) down to I_{T2}=4–6 A, which approximately follows the power law.

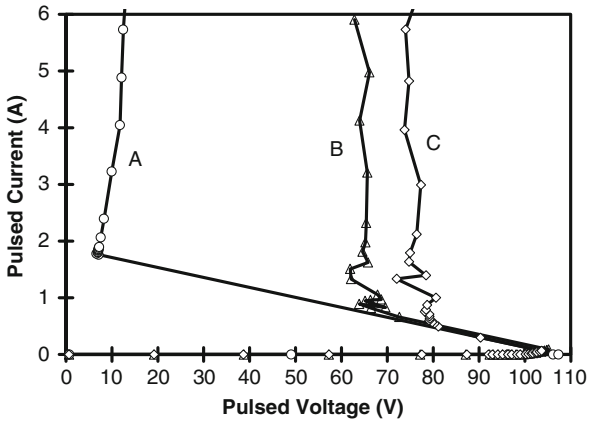


Fig. 3.50 Control of the device characteristics using n+-p+ ration and corresponding *I–V* characteristics. Device "A" – no interdigitation, solid n+ and p+ diffusions; device "B" interdigitated device with 20% of p+ area and interdigitated device "C" with 10% of p+ area

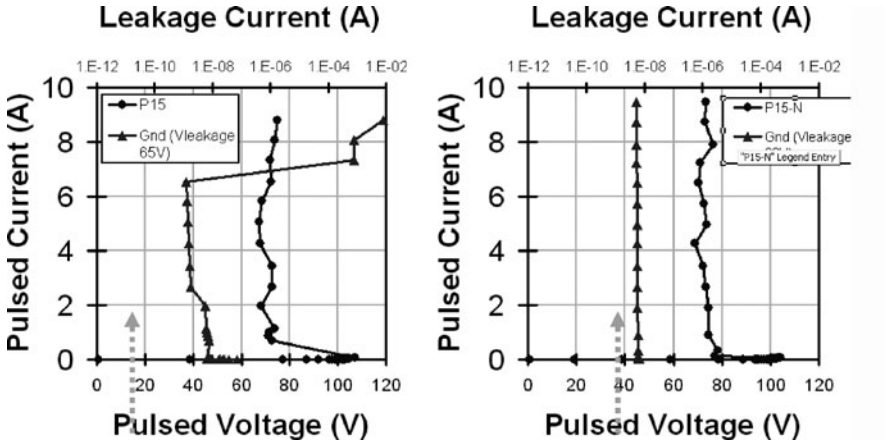


Fig. 3.51 TLP *I–V* characteristics of the dual-direction device for positive (**a**) and negative (**b**) pad current

The final level of complexity is related to the implementation of the high holding voltage characteristics for both the positive and negative currents. An example of such successful implementation is presented in Fig. 3.51. An appropriate combination of the subcollector and injection regions provides an adequate DC voltage tolerance and holding voltage for both current directions at the device PAD electrode relative to GND electrode (Fig. 3.51).

3.6.3 Dual Direction ESD Devices Based upon Si–Ge NPN BJT Structure

Methodology implementing the positive and negative feedback compensation in dual-direction device structures, described above, has limited application in CMOS or BiCMOS processes below 0.5 μm. In this case, increased well doping levels significantly reduce the gain of parasitic n–p–n and p–n–p structures. This reduction results in an inadequate CMOS-based DIAC type structure.

An example of the best DIAC structure characteristics in the 0.13 μm process is presented in Fig. 3.52. The device is implemented using the minimum spacing dimension allowed by the process rules for 0.13 μm high-frequency analog Si–Ge BiCMOS process. In comparison with the TLP characteristics of the DIAC structure implemented in a non-silicided 0.5 μm 5 V CMOS process, target performance is not reached.

In the case of the 0.13 μm process, the DIAC structures provide neither the dual-direction ESD protection nor the desired high-current tolerance.

One of the possible solutions in this case can be realized on the clamp level using isolated snapback NMOS devices. This is usually possible, since most of the scaled CMOS processes typically have an available Deep Nwell layer.

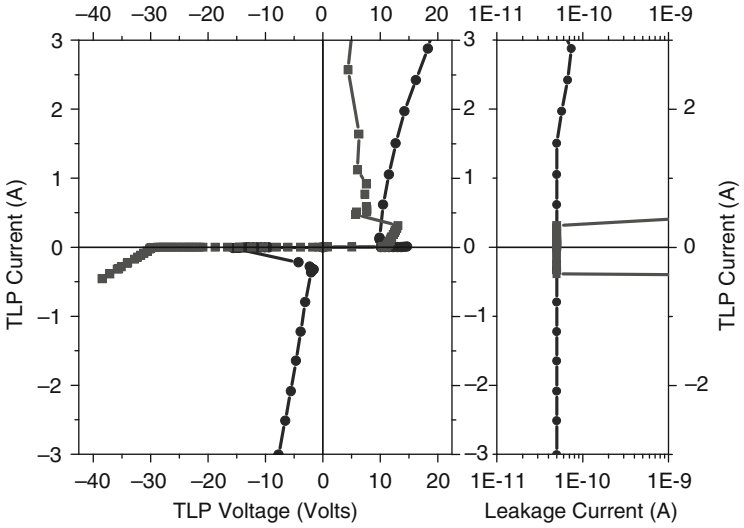


Fig. 3.52 Comparison of TLP current–voltage characteristics of the asymmetric DIAC structures (PN-n-PN) for the 0.5 μm 5 V CMOS process (*square symbols*) and 0.13 μm high-frequency analog BiCMOS process (*round symbols*)

A more aggressive alternative for small footprint solutions can be implemented in the case of BiCMOS processes, where NPN devices with a polyemitter can form a different type of the dual-direction device, as discussed below.

Similarly, at the clamp level, the dual-direction clamp can be composed of NPN BJT devices available in the process, by back-to-back stacking of the n–p–n devices (Fig. 3.53). The simple use of back-to-back isolated and stacked snapback NPN, NMOS, or BSCR devices is a straightforward approach to solve the problem of dual-direction ESD protection in standard CMOS and BiCMOS processes. This approach has the advantage of utilizing standard snapback devices and permits the option of using control electrode coupling to achieve a low triggering voltage. Nonetheless, stacking of isolated unidirectional clamps consumes an excessive space and a better solution is possible.

A more elegant device-level solution in proposed below. In order to provide scalable ESD protection, new device architecture based on a novel merged collector dual-direction BJT-based (DD-BJT) ESD clamp is proposed and successfully implemented in a 0.13 μm Si–Ge BiCMOS process.

A novel system-level solution is obtained by a merged collector DD-BJT clamp based upon SiGe NPN BJT process technology. The device architecture is based on stacked NPN BJT clamp solutions.

To achieve dual-direction ESD protection, two unidirectional NPN clamps can be placed back-to-back to produce a stacked clamp, as shown in Fig. 3.53b. The key to implementing this concept in a small footprint design is to partially eliminate and merge the subcollector regions of the stacked components. This gives

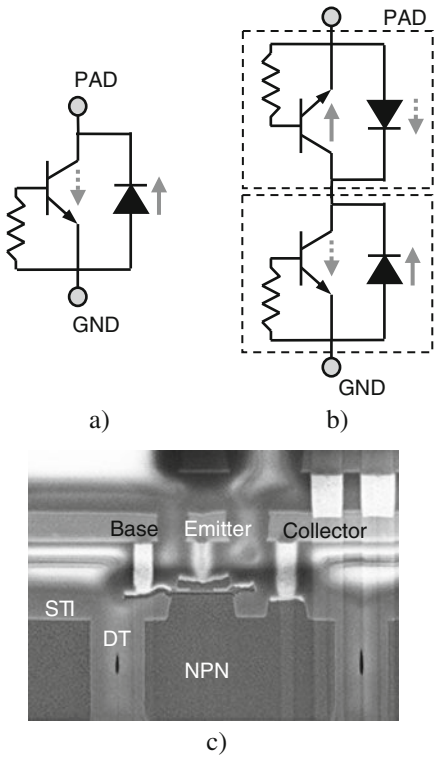


Fig. 3.53 Standard uni-direction NPN-based ESD clamp (**a**), dual-direction clamp obtained by back-to-back stacking of two uni-directional clamps (**b**) and cross-section of Si–Ge BJT with 0.24 μm emitter (**c**)

rise to a novel DD-BJT device with a cross-section and equivalent circuit shown in Fig. 3.54. At the device level, this new device is created by packing the base–emitter regions of the NPN BJT devices into one cell and removing all individual collector regions (Fig. 3.54a). This shared subcollector region approach allows for a maximum packing density for the active region of the device.

The merged collector DD-BJT clamp device has four alternating groups of fingers, represented by four circuit components in the equivalent circuit shown in Fig. 3.54c. The base–collector junction of an NPN BJT (base–emitter junction shorted) forms the reverse diode. The principle of device operation is based on the physics of the BJT NPN snapback clamp. When stressed by a positive ESD event, BJT2 remains passive and the current path is through BJT3 and the forward biased diode is formed by the base–emitter shorted BJT1. This current path is indicated in Fig. 3.54c with arrows. In the case of a negative ESD event, BJT1 is passive and the ESD current path is through BJT2 and BJT4 as depicted by the arrows.

This proposed device architecture allows for variation of both the number of fingers in each group and their arrangement. Two basic arrangements are shown in

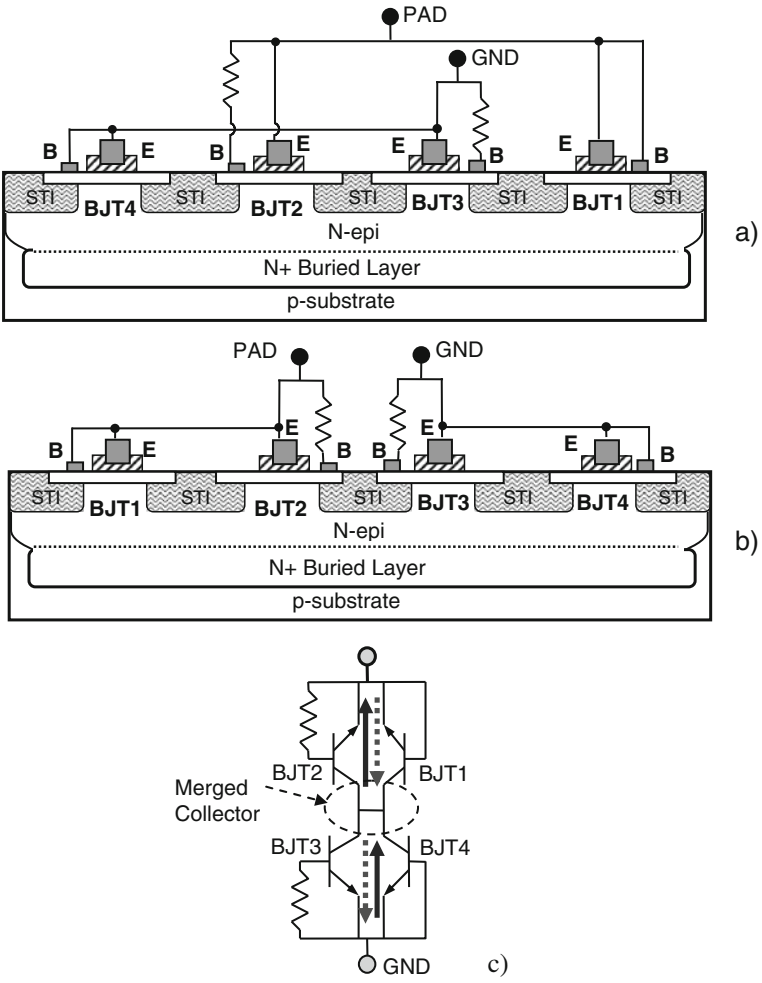


Fig. 3.54 Cross-sections (**a**, **b**) and equivalent circuit (**c**) of the merged collector DD-BJT structure

Figs. 3.54a and 3.54b. In the first case, the ESD current flows between adjacent BJT fingers; while in Fig. 3.54b, the ESD current flows between fingers separated by another BJT finger.

3.6.3.1 Experimental Characteristics of the DD-BJT Clamp

Dual-direction protection capabilities of the proposed merged collector DD-BJT ESD clamp (Fig. 3.54) have been demonstrated by standard TLP measurements. The cell under characterization had a total area of 70 μm $\times$ 70 μm. The experimental structures shown in Fig. 3.54a, b had 24 fingers (18 fingers with bases connected to

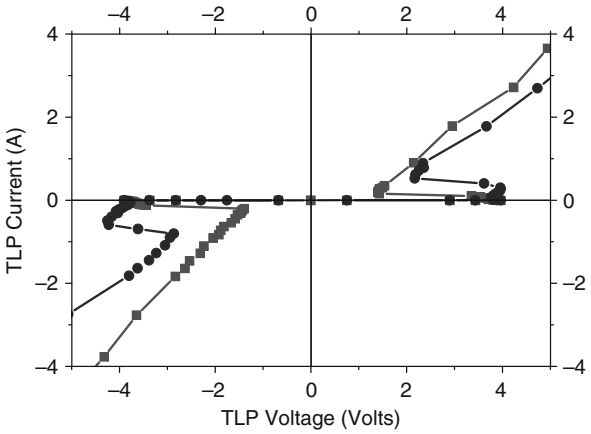


Fig. 3.55 Measured pulsed *I–V* (TLP) curve for device architectures shown in Fig. 3.54a (*square symbols*) and 3.54b (*round symbols*)

ground through a resistor and 6 fingers with a shorted base-emitter junction). Half of these fingers are connected to the pad and the other half to the ground, forming a fully symmetric dual-direction structure. Pulsed *I–V* characteristics for this device are shown in Fig. 3.55.

Both positive and negative branches of the TLP *I–V* curve of the merged collector integrated dual-direction clamp are similar to those of a unidirectional snapback NPN clamp. The triggering voltage of the dual-direction clamp is about 0.7 V higher than that of the unidirectional, due to the forward-biased diode in series with the BJT NPN in both directions. The relatively low triggering voltage is an advantage of the proposed clamp over DIAC structures, which typically have relatively a high holding voltage and a large d*V*/d*t* effect on triggering.

The holding voltage of the dual-direction clamp is lower than that of unidirectional clamp, due to the double injection (SCR effect) between the non-isolated base-emitter short and snapback NPN fingers. The holding voltage in the Fig. 3.54b configuration is higher than in that of Fig. 3.54a, because the SCR effect in the Fig. 3.54b configuration is partially suppressed by the additional separation of the fingers involved in ESD current path. However, due to the longer current path, the on-state resistance of the Fig. 3.54b configuration is larger. Additionally, the present device architecture allows control of the clamp holding voltage by varying the distance between the "diode" finger and the BJT finger and by variation of the value of the base–emitter resistor. Variation of these parameters allows the change of operation from SCR to snapback NPN mode.

For the purpose of comparison, pulsed *I–V* curves were measured for a reference unidirectional clamp based on the standard snapback NPN clamp. This clamp has 12 fingers (with collector regions) and occupies a total area of 70 μm × 50 μm. As seen in Fig. 3.56, the current tolerance of both clamps is sufficiently large,

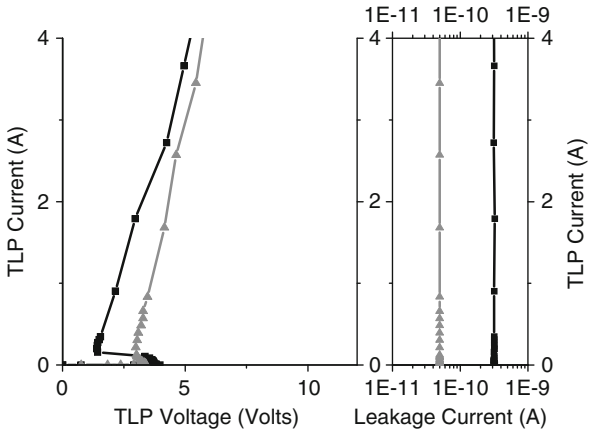


Fig. 3.56 Comparison of the positive TLP characteristic of the dual-direction merged collector integrated clamp (*square symbols*) with the prototype uni-direction BJT NPN snapback clamp (*triangle symbols*)

therefore dual-direction ESD protection does not require significantly more layout area than the unidirectional clamp. This demonstrates the advantage of the integrated dual-direction clamp in comparison with the back-to-back stacking of isolated unidirectional cells.

A valuable advantage of the suggested device architecture is the enabled device-level holding voltage control. Simple topological modifications resulting in change of the device mode operation from snapback NPN avalanche–injection conductivity mode with a high holding voltage to double injection mode in the formed SCR structure.

3.7 ESD Diodes and Passive Components

This final section presents an overview of forward-biased ESD diodes specific to analog ESD design to demonstrate common design methodology that could be applied to such components. Passive components are also discussed.

3.7.1 Forward-Biased ESD Diodes

In spite of simplicity, forward-biased ESD diodes are very important components. In the case of rail-based approach (Chapter 5), the diode supports the current path and critically contributes to the total voltage drop. In the case of high-speed I/O, the parasitics introduced by the diode structures are critical for the overall circuit performance. In the case of low-noise amplifiers (LNA) or RF pins, the ESD requirements may even have to be traded off to provide on spec chip performance.

In this section, several aspects of the operation of forward biased diodes are emphasized within the limits of this book. This is done because the description of analog ESD design will otherwise be incomplete without discussion of such a major component as an ESD diode.

There are several important considerations that should be taken into account for ESD diodes.

3.7.1.1 CMOS Diodes

In the CMOS process, the diodes are usually formed by n+ and p+ diffusions in the Nwell or Pwell (Fig. 3.57). In the case of I/O, the Pwell diode can typically be connected between the I/O pad and the GND or ESDMINUS bus, while the Nwell diodes provide a "natural" isolation from the p-substrate and is thereby suitable for connection between the I/O pad and higher supply (VDD, ESDPLUS bus) (see Chapter 5).

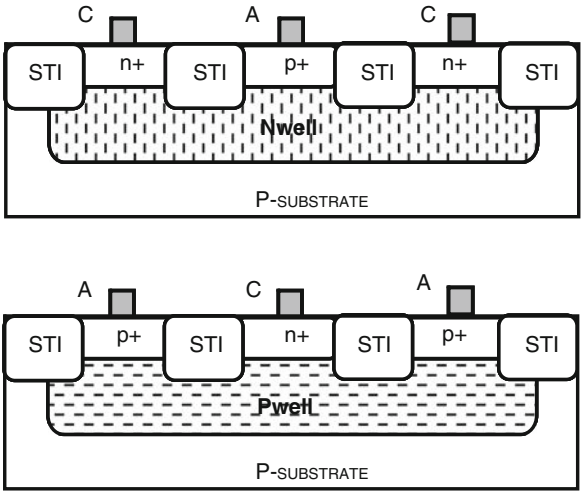


Fig. 3.57 The cross-section for Nwell and Pwell CMOS base ESD diodes

From an ESD operation point of view, it is important to emphasize that the ESD diode is acting as a lateral device at high-current level, while the parasitic capacitance and leakage are defined by the total device area.

The lower high-current resistance, lower parasitic capacitance, and lower leakage are usually desirable targets for the ESD diode spec. At first, the intuitive guess brings an idea to implement the ESD diode using a minimum diffusion length for p+ and n+ contact regions. This approach is reasonable, but not always correct.

Even in the case of non-silicided 0.5 μm processes, a more optimal diode structure may require different diode dimensions. Moreover, in the case of nanoscale process technologies, the backend, rather than the silicon, regions can already limit diode performance.

The minimum composite area (open silicon surface area without field oxide or poly) may not be the most optimal choice, due to specifics of the silicidation process. In some technologies, it can result in limited opening of the composite area due residual nitride layers. In this case, a substantial resistance may be observed. This effect may not be seen in standard devices at normal operation conditions due to a much lower current density.

Simulation results for 0.13 μm demonstrate that a minimum diffusion length is not optimal.

3.7.1.2 Gated Diodes

A practical alternative to CMOS diodes is gated diodes. The gated diodes have been proposed in [58]. Their implementation in the 0.13 μm CMOS process is shown in Fig. 3.58.

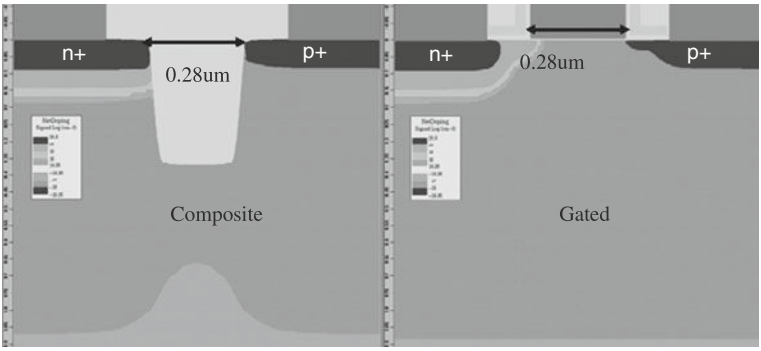


Fig. 3.58 Comparison of the process simulated cross-sections for STI (**a**) and gate (**b**) p-well diodes

These devices can provide a smaller footprint, due to lower on-state resistance and reduced parasitic capacitance of the diode clamp.

According to simulation results (Fig. 3.59), the gated diode provides a much lower (30–50%) on-state resistance, in comparison with the standard diode.

At the same time, a special effort should be applied to guarantee appropriate voltage tolerance. Skipping either nldd or pldd does not affect the forward *I–V* characteristics (Fig. 3.59). For composite diode the breakdown voltage is ~8.5 V, for gated it is ~4.5 V, and skipping nldd implant improves the breakdown voltage up to ~6.5 V (Fig. 3.60).

According to TCAD analysis data, the gated diodes are indeed a very interesting solution for aggressive high-speed I/O or system-level ESD protection. For both P- and N-Well-gated diodes with a 0.28 μm gate length, the forward current is more than 50% higher than for the corresponding composite diodes.

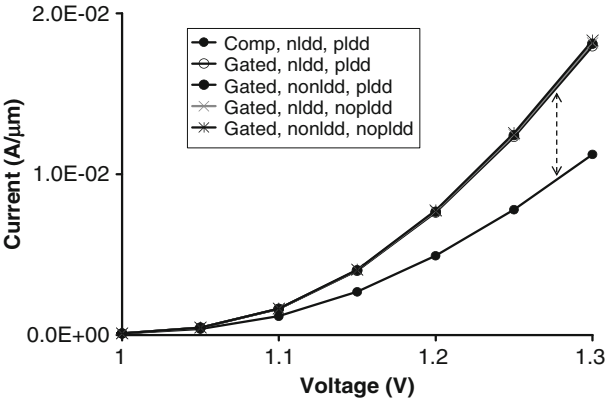


Fig. 3.59 Forward bias *I–V* characteristics for different types gated diodes implemented in 0.13 μm CMOS process with different lightly doped drain combination

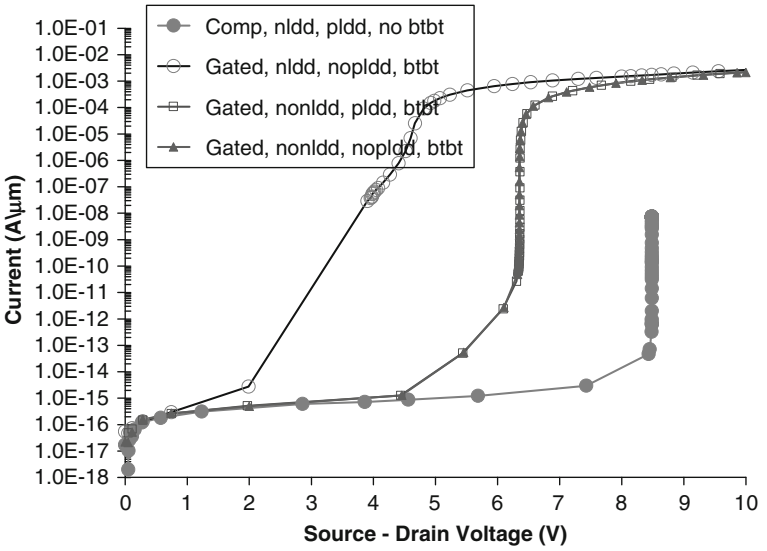


Fig. 3.60 Comparison of the reverse *I–V* characteristics of the gated diodes implemented in 0.13 μm CMOS process with different lightly doped drain combination with a regular (composite) diode

The lower breakdown voltage of gated diodes is a concern, however, skipping hnldd implant for P- and pldd implant for N-Well diodes leaves the forward current unaffected while increasing the breakdown voltage by at least 2 V.

3.7.2 Passives

The robustness of passive components is also critical for analog ESD design. Poly and diffusion resistors are often used to separate different inputs and control pins

from the pad, implementing two-stage ESD protection (Chapter 4). Thin film resistors might interface with the pad directly, for example, in precision amplifiers. MIM and poly capacitors may be part of both the ESD clamp and the analog circuit components that interface with chip pins. This subsection highlights some aspects of passive components performance specific to ESD conditions.

3.7.2.1 Saturation Resistors [59]

A saturation resistor can be formed using a layer available in the process both at the surface (n+, nldd, pldd, Pwell, Nwell, etc.) and in the depth of the device by reusing subcollector regions.

The resistor provides a close to linear *I–V* dependence up to the saturation current level, where the density of the injected carriers becomes comparable with the doping level. After the saturation region, the conductivity modulation phenomenon results in device snapback. This effect was already discussed in Chapter 2 for n+–n–n+ structures.

For saturation resistors and many other passive components, TLP or dc curve trace characterization is required to determine the SOA of the devices with specific parameters.

Diffusion resistors should be always assessed according to their voltage tolerance. Other undesirable features are the parasitic capacitive coupling to substrate and noise, besides the limited resistance range and the additional space required on the chip. That is why preference is often given to unsilicided poly resistors as devices of choice for ESD clamps and the ESD network.

Some experimental data for P-buried layer (PBL) and N-buried layer (NBL) resistors are presented below. In the process, an n-epi is grown after the N-buried layer. The doping in the NBL is approximately 3×10^{19} cm^{-3}, compared to the 5×10^{15} cm^{-3} doping level in the epi layer. The buried layer is contacted through an n-type poly and N-sinker. The resistor is completely isolated from the substrate and adjacent devices, due to the deep trench and bonded substrate (SOI). Figure 3.61c is a 1D doping profile from the process simulation of the NPN transistor and shows the doping profile of the Nepi, N-sinker, and NBL layers. Figure 3.61a shows a 2D cross-section schematic of the NBL resistor.

The PBL forms the PBL resistor. The Nepi, which is grown after the buried layer, is counter-doped using the Pwell. The doping profiles are shown in the 1D cross-section of the PNP transistor, obtained from a process simulation. The 2D cross-section schematic in Fig. 3.61b illustrates the structure of the PBL resistor.

A typical layout of a buried resistor is shown in Fig. 3.62. The *L* and *W* indicated on the drawing correspond to the length and width of the resistor in the following sections. Three rows of contacts have been drawn on both sides, with a maximum number of contacts for the width. In the 10 μm-wide resistor, there are three rows of three contacts, and in the 20 μm wide resistor, there are three rows of eight contacts.

The results of the curve trace indicate that there is not a linear scaling of resistance with squares, as defined using the drawn *W* and *L* parameters from the layout (Fig. 3.63). There is an additional resistance associated with the sinker, which has not been accounted for. The scaling of the resistance with a number of drawn squares

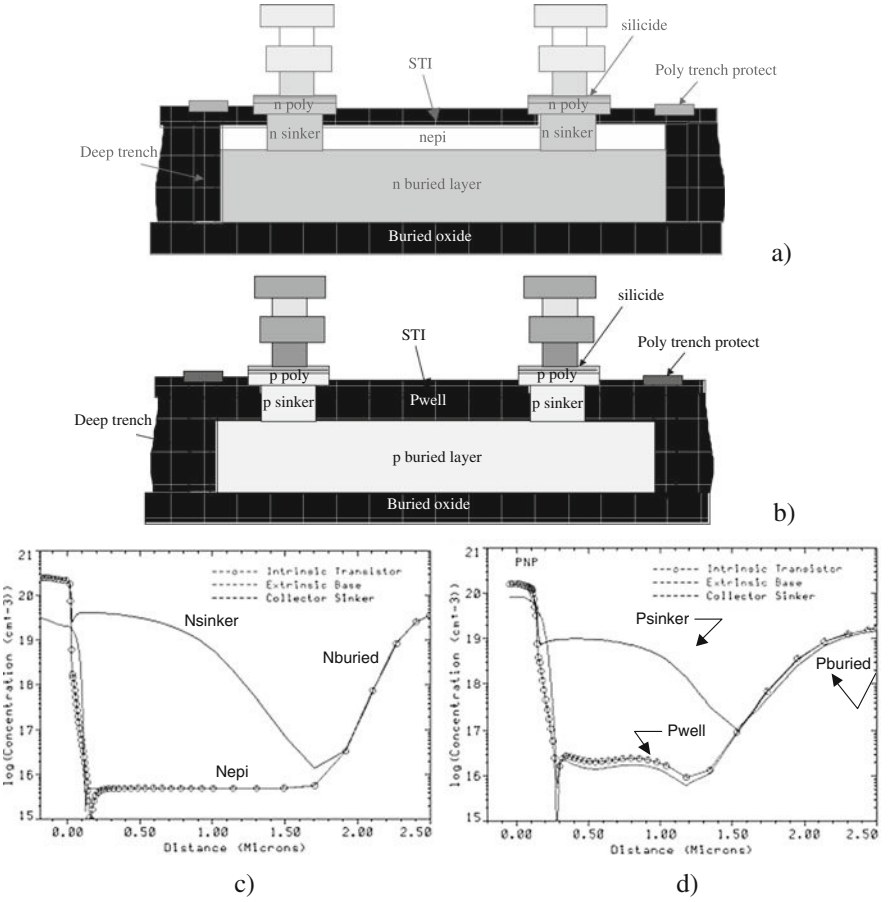


Fig. 3.61 Two-dimensional cross-section schematic illustrating the construction of the NBL (**a**) and PBL (**b**) resistors, and 1D doping profile of the NPN (**c**) and PNP (**d**) transistors from the process simulation

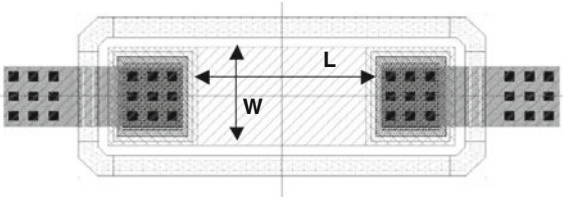


Fig. 3.62 Typical layout of the resistor

clearly illustrates that the resistance cannot be based purely on the buried layer between the sinkers. In fact, the drawn resistor at 2.5 squares has a lower resistance than the resistor drawn at 2 squares. Thus, in general, implementation of saturation resistors requires some characterization effort.

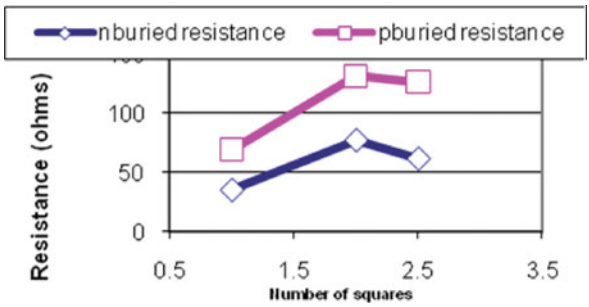


Fig. 3.63 Scaling of resistance with number of squares, as defined at layout

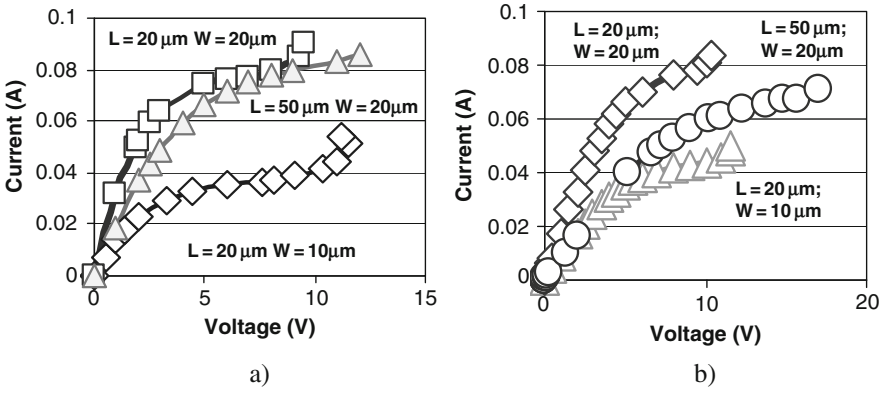


Fig. 3.64 *I–V* characteristics of the NBL (**a**) and PBL (**b**) resistors measured on the curve tracer

Breakdown voltage for the buried resistors varied between 10 and 12 V, with the exception of the PBL *L*50 × *W*20 resistor that exhibited a higher breakdown voltage at 17.5 V (Fig. 3.64). All the resistors measured exhibited a snapback-type of characteristic, but this was not recorded in the curve trace, due to the destructive nature of this measurement. The resistors provide a region of current saturation above 2–3 V in the N-buried resistors, and above 5 V in the PBL resistors.

The TLP curves measured on the n-buried resistors show a saturation region between 5 and 12 V, but then the saturation region appears to collapse and a region of lower resistance dominates (Fig. 3.65). However, in the p-buried resistors, no saturation region is observed up to 20 V (Fig. 3.65).

The resistance is not defined solely by the geometry of the buried layer next to the sinker. In addition to the contribution of the contact resistance, the doping profiles of the sinker and buried layer must be accounted for in the calculation of resistance scaling. This is a good topic for TCAD analysis and may require 3D analysis.

The second observation shows that the resistors do not behave as saturation resistors at ESD conditions. Using the resistance extracted from the TLP measurements,

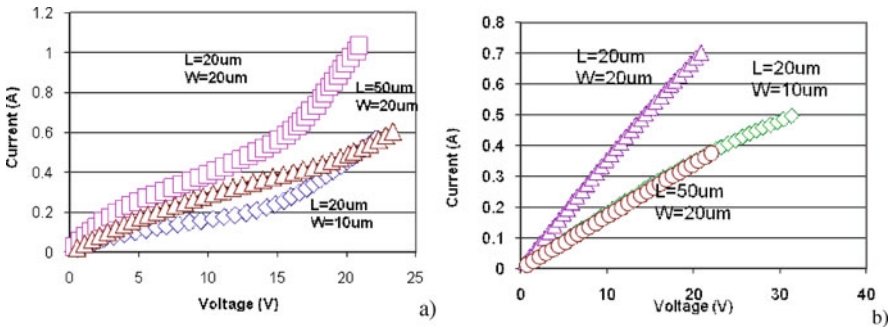


Fig. 3.65 *I–V* characteristics of the n-buried (**a**) and p-buried (**b**) resistors measured on the TLP system at ESD conditions

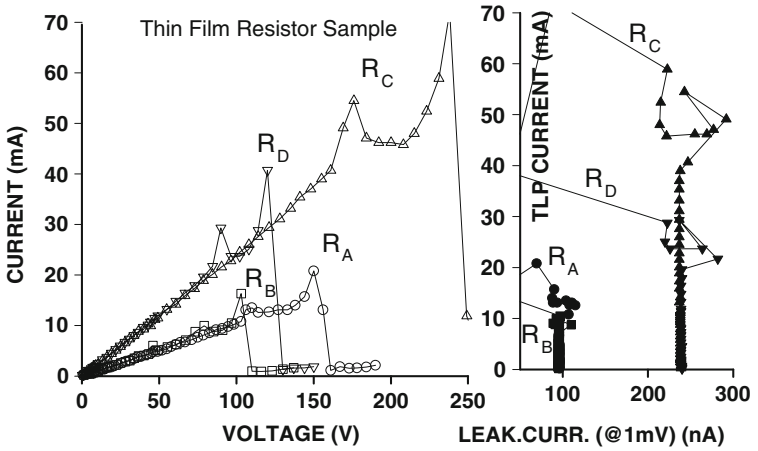


Fig. 3.66 TLP characteristics of the selected resistor structures with parameters indicated in Table 2.1

it should be possible to use the resistors to limit current, if the element protected can sustain the voltage drop.

3.7.2.2 Thin Film Resistors

Thin film resistors (TFR) are used in many analog applications, for example, in precision amplifiers.

An example of ESD *I–V* characteristics for a thin film resistor with a sheet resistance of ~1 kΩ/sq is presented below. A SiCCr film with precision matching (<0.1% at 3 sigma deviation) forms the resistor.

The maximum ESD current depends on the aspect ratio of the TFR. Resistor width plays the main role in the resistor's ability to handle high-current values: the

wider the resistor, the larger current that it can handle. Even if inputs are separated by large value of TFR, ESD clamps are always necessary, at least to protect the pad dielectric. Typically, pad dielectric can withstand a voltage of ~200 V, which is significantly lower than the voltage amplitude during the ESD event, even in the package-level test. At the same time, ESD protection may be needed for the resistor itself, if under the ESD stress there is residual current flowing through the resistor, as the ESD clamp keeps it biased at a certain voltage value.

Normal operating DC current density (per width) for these resistors is 0.05 mA/um. Typically, additional ESD protection for the resistor is deemed necessary if the residual current value under ESD stress is 40× of the operating current density, i.e., 2 mA/μm.

Maximal ESD current density is ~3 mA/μm. The maximal operating current is $40 \times 0.05 = 2$ mA/um, which is very close to the max ESD value. For conservative ESD design, this resistor has to be either made very wide or have an additional clamp that fixes current at <2 mA/um. Otherwise, further reliability studies are needed to more precisely find the boundary of the resistor robustness, i.e., neglect the 40× rule and find the exact max operating DC current.

Both the critical power and the absolute maximum pulsed voltage of the resistor depend upon the total resistor area. The power and voltage increases with increase of the area at a given resistor value. Within measurement accuracy, the specific critical power for the beginning of degradation is ~10 mW per square micron (Table 3.1).

Table 3.1 Summary of the experimental data for TFR

Parameter	Sample R_A	Sample R_B	Sample R_C	Sample R_D
Resistor value (kΩ)	10	10	4	4
Total area (width × length μm^2)	3 × 30	2 × 20	12 × 48	6 × 24
Critical snapback voltage (V)	98.4	61.2	155	78.6
Critical snapback current (mA)	10.2	6.2	39	19.6
Critical pulsed power for degradation (W)	1	3.8	6	1.54
Critical current density per resistor width (mA/μm)	3.4	3.1	3.3	3.3
Specific critical power for degradation (W/μm2)	11.2	9.5	10.5	10.7

3.8 Summary

In this chapter, it is demonstrated that the phenomena of the same nature are responsible for both limiting the absolute maximum voltage and determining critical snapback voltages in standard devices. In each case, the snapback effect in the device can be reduced to fundamental conductivity modulation mechanisms that provide structure-level positive feedback.

The reversible operation of ESD devices in pulse conditions is achieved by implementation of the device-level negative feedback to limit current increase caused by positive feedback. Both approaches are suitable for engineering ESD devices with appropriate characteristics for optimal voltage waveform at ESD pulse. An innovative use of device-level negative feedback provides an opportunity to create self-protected devices.

A set of transformation rules is demonstrated to enable implementation of the ESD devices by modification of the standard devices. The focus of the ESD device material further illustrates that in most cases, the problem of ESD protection can be solved at the level of free ESD device implementation, which avoids alteration of the process technology or implementation of additional mask layers. Specific efforts are applied to design devices with low sensitivity to mask misalignment. ESD device engineering requires a strong understanding of integrated process technologies architecture and process-step specifics, while also requiring substantial experimental verification and validation of the final solutions.

At the same time, the ESD device can contain a number of parasitic structures that provide alternative current path scenarios, internal reference, and components for reverse path ESD current.

The physics of operation of the devices discussed in this chapter can be generally understood through the analysis of the 2D cross-sections. Thus, numerical simulation tools are adequately accurate for comparative analysis and validation of the new ideas.

The next chapter brings the ESD devices into 2.5D or 3D levels, where additional dimension is involved to treat the ESD device as a distributed integrated object fabricated on the surface of the semiconductor die. This aspect is one of the features that separate the ESD device level from the clamp level. Another feature involved in the clamp level is the subcircuit, which can be attached to the control electrode of ESD device. The subcircuit is used to control the dc, triggering, and holding voltage characteristics. Thus, the characteristics of ESD devices at different control electrode conditions are a rather important part of the easily collected experimental results. In addition the next chapter will demonstrate the significant advantage that can be obtained from an active control of the triggering characteristics of the ESD device by a driver.

DECIMM^{TM} Simulation Examples for Chapter 3

To download a trial version of the numerical simulation software and request an electronic license key please visit http://www.analogesd.com

To download libraries with simulation examples for this chapter please visit http://www.analogesd.com/Chapter2.html

List of examples is subject to change.

Example 3.1 Standard Devices in BCD Process Technology

Library Name: Examples3 Default BCD and ESD Devices
Project Names: E3.1a_HNMOS; E3.1b_HPMOS; E3.1c_NPN; E3.1d_PNP

This set of examples provides fully parameterized device structures representing devices most commonly used in typical 0.5 μm BCD process technology. Among library devices, there are 5 V NMOS (Fig. E3.1a), 5 V PMOS (Fig. E3.1b), N-well and P-well diodes (Fig. E3.1c), 24 V NPN and PNP BJT devices (Fig. E3.1d),

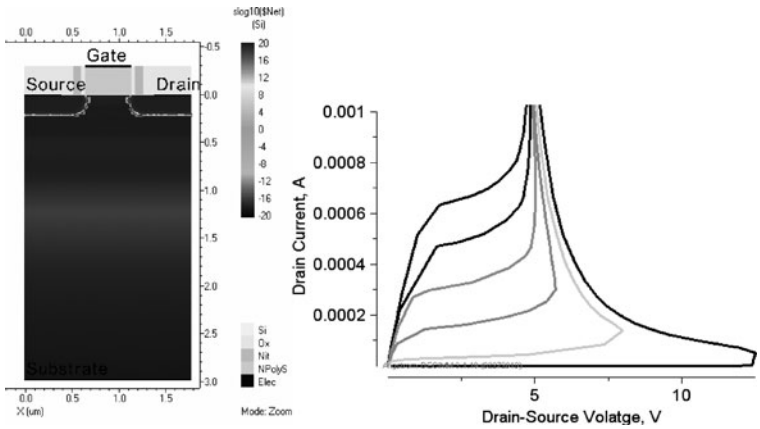


Fig. E3.1a Cross-section of fully parameterized NMOS structure and simulated family of drain curves for different gate bias values

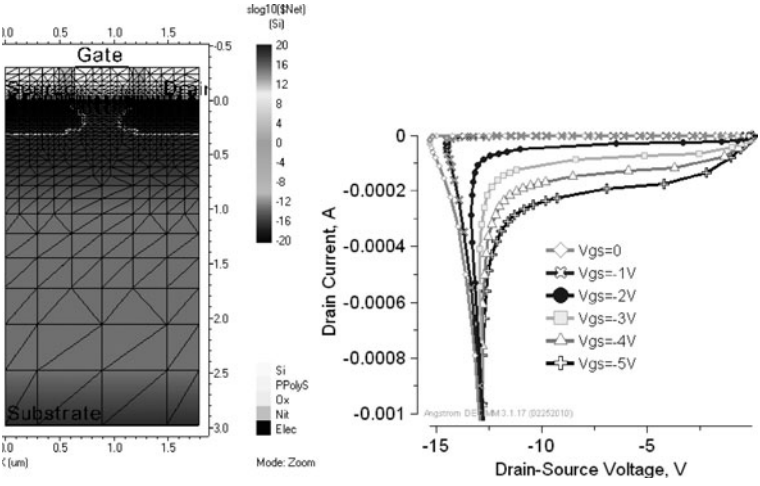


Fig. E3.1b Cross-section and simulated output simulation results for the default 5 V PMOS device

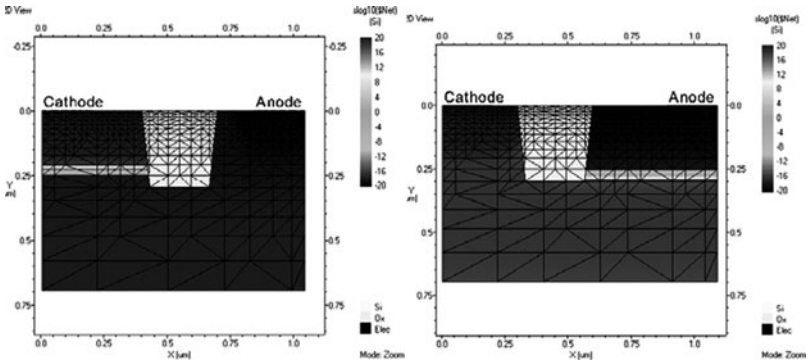


Fig. E3.1c View of fully parameterized cross-section for N-well and P-well diode structures compatible with the CMOS process

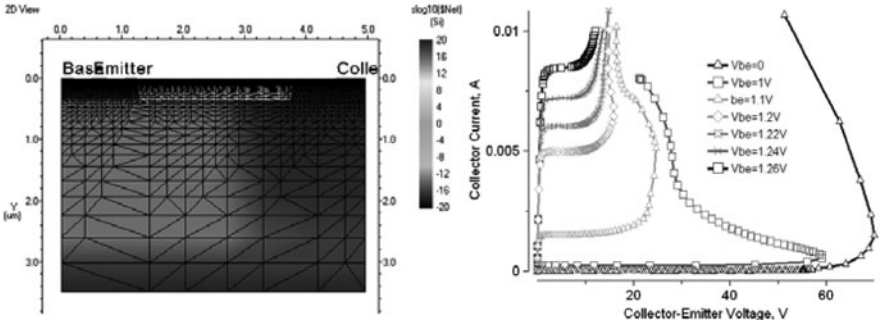


Fig. E3.1d Cross-section for fully parameterized 24 V NPN BJT and family of simulated output characteristics

24 V NLDMOS (Fig. 3.1e). In addition, vertical trench DMOS and IGBT devices are also included (Fig. 3.1f).

The main goal of the numerical analysis presented here is simulation of the quasi-static isothermal *I–V* characteristics, including the regimes of avalanche breakdown, strong injection, and conductivity modulation. Utilization of fully parameterized device structure geometries and implant profiles makes it easy to explore the impact of changes in device structure on the device performance in different regimes.

Example 3.2 Typical ESD Devices in 0.5 μm BCD Process Technology

Library Name: Examples3 Default BCD and ESD Devices *Project Names*: E3.2a_SNMOS; E3.2b_NLVTSCR; E3.2c_PLVTSCR; E3.2d_FOXSCR; E3.2e_BSCR; E3.2 g_NLDMOS_SCR

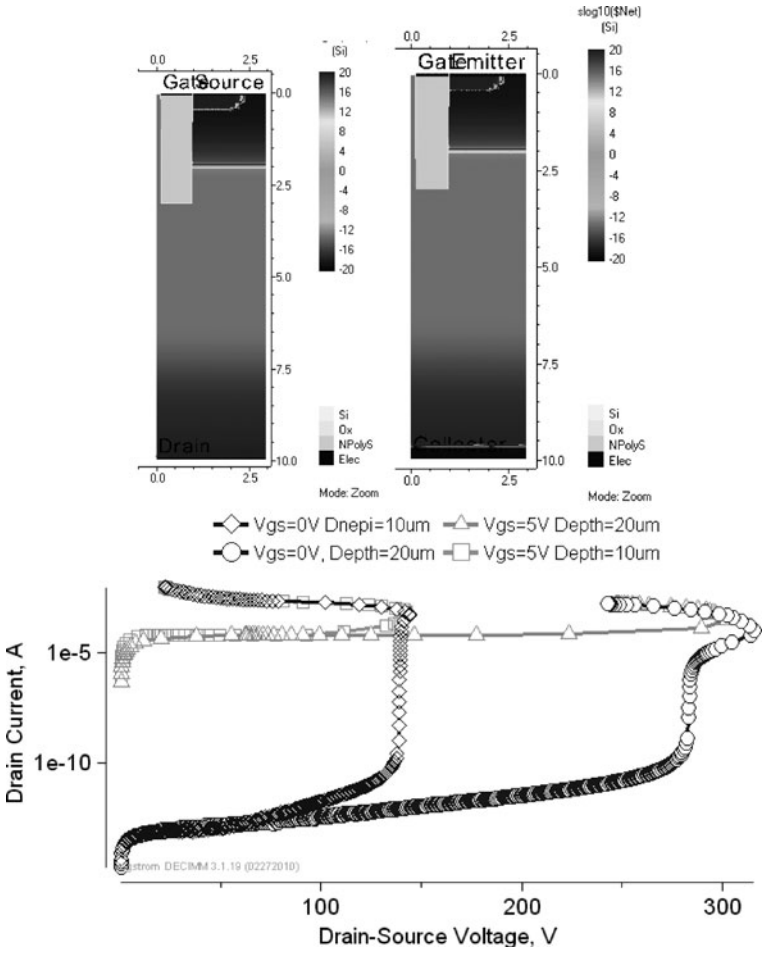


Fig. E3.1e Cross-section for fully parameterized 100–600 V vertical DMOS and IGBT structures and simulated isothermal drain *I–V* characteristics for vertical DMOS for different N-epi depth values and gate bias conditions

This set of examples provides fully parameterized device structures representing typical snapback ESD devices most commonly used in local ESD protection clamps in typical 0.5 μm BCD process technology. Among many library devices, there are 5 V Snapback NMOS (Fig. E3.2a), high-voltage bipolar SCR and CMOS based field-oxide SCR (Fig. E3.2b), 5 V NLVTSCR, 5 V PLVTSCR, and 24 V NLDMOS SCR (Fig. E3.2c). Suggested use of the numerical analysis includes simulation of the quasi-static isothermal *I–V* characteristics, breakdown and holding voltage and on-state resistance as a function of the device structure parameters. Utilization of

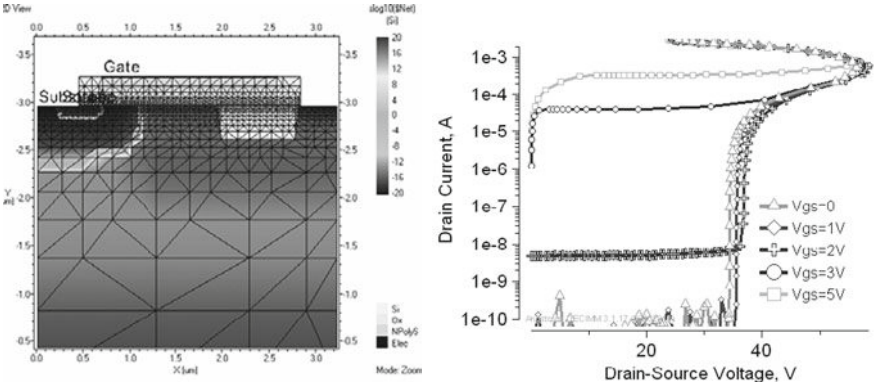


Fig. E3.1f Cross-section and simulated output drain *I–V* characteristics for fully parameterized the default 20 V NLDMOS device

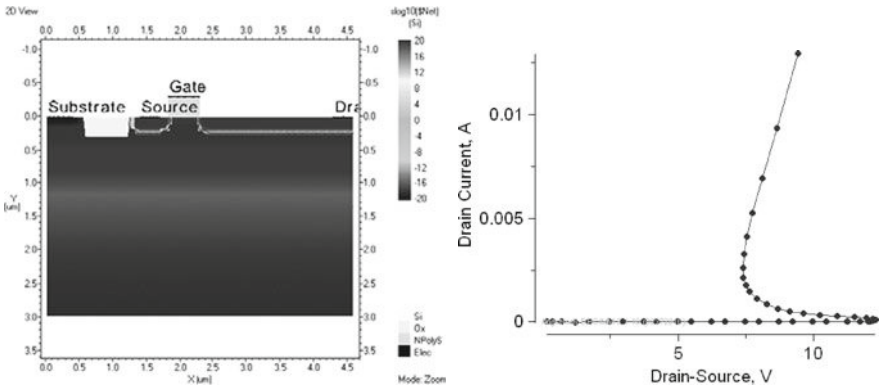


Fig. E3.2a Cross-section of fully parameterized 5 V snapback NMOS structure and simulated output *I–V* characteristic

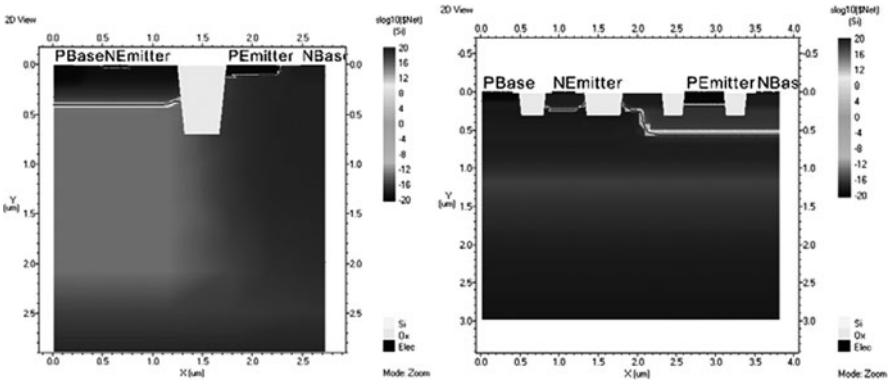


Fig. E3.2b Cross-section of fully parameterized high-voltage bipolar SCR and field oxide SCR

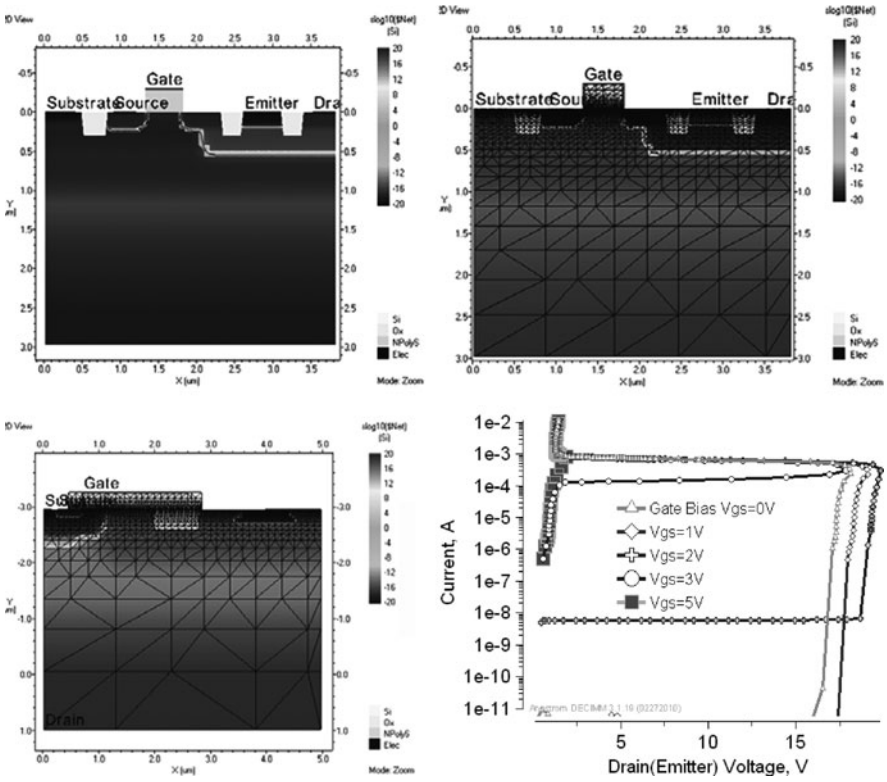


Fig. E3.2c Cross-section of fully parameterized 5 V tolerant NLVTSCR and PLVTSCR and high-voltage NLDMOS-SCR with simulated *I–V* characteristics for 15 V version

fully parameterized device structure geometries and implant profiles facilitates study of the impact of changes in device structure on the device operation in different regimes.

Example 3.3 Ring Oscillators

Library Name: Examples3 Default BCD and ESD Devices*Project Name*: E3.3a_Ring_Oscillator_w_1V5_Compact_Model_Devices*Project Name*: E3.3b_Ring_Oscillator_w_FE_5 V_CMOS_Devices

These examples illustrate operation of ring oscillator circuits and highlight versatility of the mixed-mode simulation approach, which allows employment of both

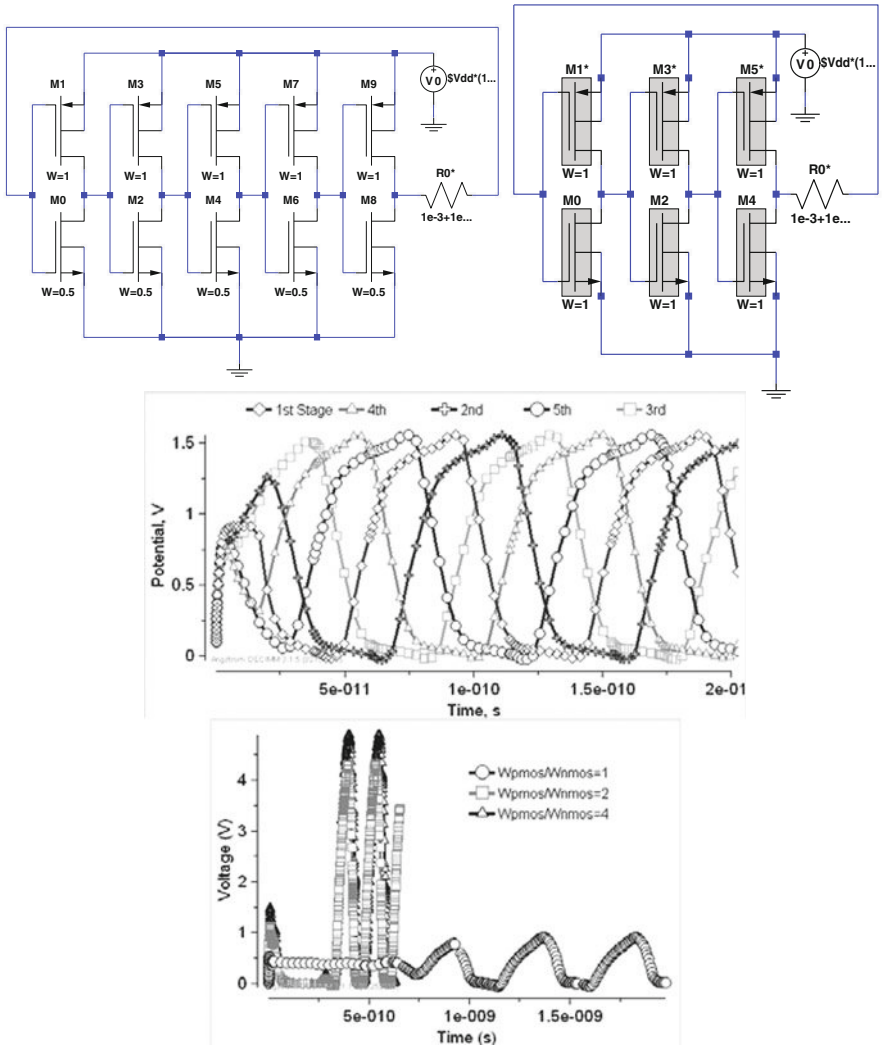


Fig. E3.3 Mixed-mode simulation circuits for five- and three-stage ring oscillators built using compact models and FEM devices, respectively. Voltage waveforms at different stages of compact RO, and different PMOS/NMOS width ratios in the FEM RO circuit

highly physically accurate FEM devices and computationally efficient compact models to achieve high simulation speed and accuracy.

Chapter 4
ESD Clamps

The previous chapter described ESD devices with emphasis on the positive and negative feedback effects realized on the structure level of the devices. The positive feedback effects have been identified through the conductivity modulation phenomena in the parasitic n–p–n, p–n–p, or p–n–p–n elementary structures. At the same time, ESD devices already include local and non-local current limiting on the device level. This feature provides the negative feedback that is used to suppress the excessive positive feedback and limit uncontrollable current density increase. Thus, the ESD device structure combines the active device regions with blocking junctions, the drift region, RESURF regions, control electrodes, and the contacts.

At the same time, a real implementation of the ESD protection device in the ESD protection cell for integrated process technology involves many more different aspects. These include device width scaling, lateral isolation, vertical isolation from substrate, latch-up guard rings and, of course, all the clamp components for voltage reference and dynamic coupling.

As an object, the ESD clamp (or ESD cell) presents the next level of hierarchy for ESD protection solution engineering in semiconductor integrated process technology. Often, the ESD clamps are treated simply as analog circuit building blocks. Thus, intellectual property (IP) for the clamps is usually released in the form of ESD libraries for the given CAD environment. Clamp IP development is usually completed according to a specification for the envisioned product pins spec in correlation with ESD protection network design.

In general, a clamp combines a distributed ESD device with appropriate width scaling according to the ESD current specification and additional circuit components. The additional circuit components are optional and are used to bring the triggering characteristics of the device within the ESD protection window. Often, for the ESD device with internal voltage reference, the clamp component count can be relatively small, containing, for example, only the reverse path diode. However, in some cases demonstrated in this chapter, rather complex drivers can be used to enable and disable the ESD clamp depending on operation conditions.

This means, in principle, that the clamp can be designed based upon a single ESD device structure, if the ESD device itself provides appropriate characteristics that already meet the protection clamp specifications. The following parameters of the clamp are usually major figures of merit: the voltage triggering the device into a

V.A. Vashchenko, A. Shibkov, *ESD Design for Analog Circuits*,
DOI 10.1007/978-1-4419-6565-3_4, © Springer Science+Business Media, LLC 2010

high-current state, the holding voltage, the transient response (turn-on time, voltage waveform), dc voltage tolerance, the breakdown voltage at a given current, the leakage current for the maximum operation voltage, dimensions of the clamp footprint, used mask layers count for active layers and metallization, RF parasitic, reverse current clamping, recovery time, and some other parameters more specific to the particular application.

In most cases of package level requirements, the triggering characteristics of the clamps should be guaranteed within normal operation temperature conditions, while the clamps' parasitic characteristics should be provided for the whole operational temperature range. In the case of system level ESD protection, the specification may include much broader requirements.

Often, the clamp requires additional components to incorporate all the necessary capabilities.

The most typical clamp components include the following:

(i) Triggering circuit: a dynamic coupling, reference voltage, or reference current sub-circuit. This subcircuit is used to provide clamp turn-on inside the ESD protection window voltage range.
(ii) Holding voltage control components.
(iii) High-current ballasting and current crowding prevention design.
(iv) Reverse path diode for voltage clamping at negative current direction when the pad is biased at negative voltage.
(v) Latch-up isolation guard rings.
(vi) Secondary and following stages components.

The clamp may also include rather sophisticated subcircuits with control, shutdown, or enable functions selected depending upon the internal circuit state. At the same time, one of the trends in creating small footprint solutions is the use of internal device-level features that integrate the clamp components inside the device structure.

The other major difference between the ESD clamp and the ESD device structure is the specific clamp layout design for both drawn and generated process mask layers. This difference is essentially between the 2D and 3D ESD devices, or at least the 2.5D physical description required for the clamp.

Layout implementation of the ESD device is a very critical part of the clamp. It brings a number of important peculiarities. Often, minor changes on the layout level may significantly change the triggering and other characteristics of the final clamp. This is why a basic principle of ESD design is verification of the re-designed clamp characteristics on the test chip prior to application of the clamp in the product.

Usually, ESD protection clamps are placed near the bond pads or micro-SMD (surface mounting device) bumps to avoid the excessive voltage drop on the metallization routing and to reduce the impact of the injection current in the clamp on the internal circuit latch-up. Thus, the pad and pitch dimensions across the pad ring usually determine the width of the ESD clamp that supports a product-independent ESD pad ring design. In most cases, the clamp width in the layout is a specified

parameter. For example, at a pad and pitch spacing of ~120 μm, the drawn clamp layout width should not exceed this size. At the same time, in order to collect a ~1.5 A current from an ESD device with a current density of ~3 mA/μm, the required total device width is ~300 μm. Thus, in most cases, the ESD device should be designed in the form of an array. The total amount of fingers N multiplied by the finger width w in the array is usually within the range $W=N\times w=100–1000$ μm.

In such a multifinger distributed ESD device, it is either hard or simply inefficient to make each part of the device with identical internal parameters. Due to truly 3D design, each different part of the finger and the different fingers themselves may have a different metallization resistance, different effective spacing to the blocking junction connection, or different control electrode reference.

Thus, the device cross-sections for each segment in the distributed layout can only roughly be considered congruent. For example, in a snapback NMOS cell (Section 4.2), the distance between the well tap and each finger's source diffusion region is variable. This results in a difference in the effective base resistance of the parasitic n–p–n structure. As a result, the cell turn-on can hardly be expected to be uniform. Moreover, perhaps the most typical scenario is non-uniformity of the middle or corner of the distributed ESD device in the cell.

Nevertheless, it is usually expected that in a properly designed ESD cell, each micron of its width can contribute to the total ESD device current. In other words, linear width scaling is assumed upon device width. This is possible because of the propagation of conductivity modulation across the cell during turn-on, due to mutual coupling of the device segments.

This simple understanding brings another important principle of ESD cell design. The principle requires implementation of all the fingers of the ESD device within the same shared well, epi, or other body layer of the device. This is done specifically to enable sharing between the distributed device segments in opposite to, for example, composing the cell by paralleling fully isolated fingers to collect the appropriate level of current. Not following this principle can create a passing current level issue due to multifinger turn-on [30] or cause very non-linear width scaling.

On the other hand, the topological degree of freedom is often advantageous in engineering appropriate characteristics of the clamp itself. Thus, the topology of ESD devices inside the clamps is the first aspect addressed in this chapter.

The second major subject is the reference subcircuits and techniques that can be used to implement appropriate clamp characteristics during both ESD operation and under normal operation conditions.

There is a tremendous variety of different ESD clamps and solution variations covered by numerous original papers in the ESD field. The major source of up-to-date information is the on-chip and device physics sections in the Proceedings of the EOSESD Symposium [8]. One of the most recent and complete reviews for different solutions is presented in [4–6].

The purpose of this chapter, however, is not to present most of the available findings in the ESD clamp design field. Instead, similarly to all the following chapters, only the most practical examples are used for in-depth discussion of the major principles and methodologies involved in design.

This is done due to the following two reasons. First of all, this information is limited in order to avoid overloading the book with excessive number of examples. The second reason consists in the fact that in most cases, practical implementation of ESD protection solutions requires experimentation and verification specific to the particular process technology.

Based upon the authors' experience, it is rather rare that solutions directly based on publication can be expected to provide target performance for the new process. The printed space of this book is used to provide understanding of ESD clamp design rather than to serve as an encyclopedia or reference manual for all possible clamp implementations. Such an approach is expected to be more useful for practical ESD design of clamps in new process technologies.

The material in the chapter will be presented in two steps. First, the most typical ESD protection clamps are discussed. Then, several more complex examples are discussed to illustrate the potential for innovation in the ESD design field. This basic material is followed by more advanced sections focused on complex clamp subcircuits, self-protection, and system level protection. The final classification of the solutions is presented at the end of the section.

4.1 Active NMOS Clamp

Despite the focus of this book being on clamps based upon ESD devices operating in conductivity modulation regimes, this section discusses the alternative solution first. This alternative solution is the so-called active clamp that does not operate in snapback. The first such clamp solution has been proposed in [60], and since then became the most widely used solution for high pin count products, as well as digital and low-voltage domains. There are many variations of implementations of the active NMOS and PMOS clamps [61–70] that are used by industry leaders (see Chapter 5 too).

Originally, the major principle of such clamp design was the rise-time-dependent turn-on of the big distributed NMOS device in normal current conduction mode. The dependence upon rise time provides turn-on during ESD pulse, but suppresses clamp turn-on during a much slower powering voltage ramp.

In spite of certain application-related limitations, this clamp-level solution should be considered, perhaps, first in any new design.

The major advantage of this clamp is that it does not require ESD snapback devices. The solution enables a complete ESD waveform simulation using compact models for the clamp components which are usually available.

A typical circuit diagram and layout of the clamp cells is presented in Fig. 4.1. The clamp is composed of an RC-timer cell and a number of "slave" clamps. The slave clamps contain an NMOS device with a low on-state saturation resistance that provides monopolar channel conductivity during ESD operation.

In a simple case, the RC-timer is designed for typical time constant of ~4–8 μs. This guarantees a low current state of the slave clamp NMOS device during possible transient in normal operation conditions and provides relatively low parasitic current

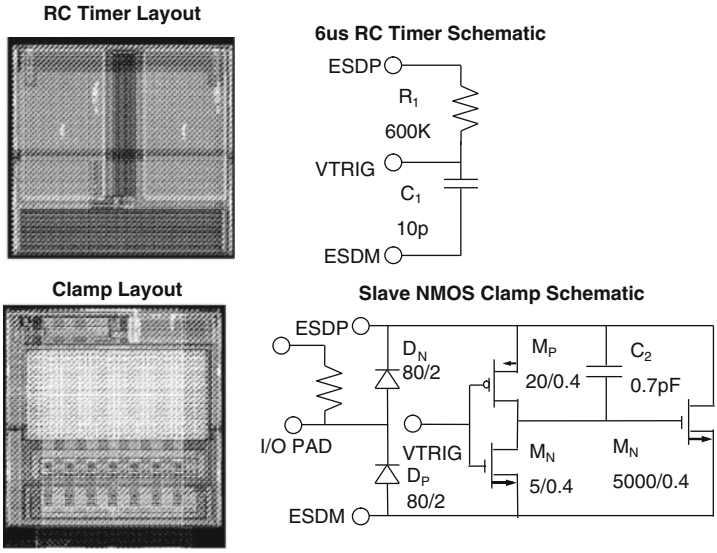


Fig. 4.1 Circuit diagram for the active clamp with RC-timer (**a**) and slave clamp cells (**b**)

during the slow powering sequence. At the same time, the time constant is sufficient to completely discharge the HBM ESD pulse to safe voltage levels with a decay time of 150 ns.

The clamp has two distinctive regimes of operation: the normal operation regime and the ESD operation regime.

In the case of ESD event, fast transient voltage at the pad pin results in a corresponding rising voltage on the drain of the slave NMOS devices. Due to the drain-gate capacitive coupling, the transient gate bias of the NMOS device exceeds the threshold voltage which switches it into the "on" state with low channel resistance. To improve the coupling voltage, speed-up capacitance C_N is used in every slave clamp. Due to the on-state condition, an NMOS device with an appropriate total width provides a discharge current path with a low-voltage drop. In a properly designed clamp, this results in a voltage limitation below the 1–2 V level. After the first ~300 ns of fast transient operation, the slave clamp inverters regain control of the NMOS device's gate potential which is now determined by the RC-timer state. The RC time constant for the RC-timer is usually chosen ~6 μs. After the RC timer capacitor C_T is sufficiently charged through the resistor R_T, the slave inverter switches the NMOS device into a low leakage "off" state.

In the case of a normal powering up sequence, the power supply voltage ramp is usually relatively slow. This results in a rather small parasitic current through the clamp due to the relatively fast response of the RC-timer.

A relatively large width NMOS device of ~4–10 mm is required to support an ESD current level. The RC-timer components are space consuming as well since

large area capacitor and resistor are required. The ESD protection network (Chapter 5) includes several slave clamps composed with fewer RC-timers.

The implementation of active clamps in the ESD protection network will be discussed in Section 5.1.

A more complex approach is proposed in [71]. This proposed circuit fulfills a dual function by detecting the slew rate and controlling the duration of the turn-on time of the active clamp. Detection of the slew rate has the advantage of recognizing input pulses as ESD or other, such as hot plug-in or power on.

There are several limitations related to the active clamp design:

(i) Clamp size: usually, ESD protection with clamps becomes efficient in analog circuits only in the case of 8–10 pins that belong to the same voltage domains, for example, in a digital interface.
(ii) Low-voltage NMOS: the clamp has a reasonable size only in case of low-voltage NMOS devices. Clamp design using NLDMOS is in most cases impractical. Similarly, the voltage tolerance of the clamp is limited by the hot carrier reliability limitations on the drain-gate bias of low-voltage NMOS.
(iii) System-level and high ESD level spec solutions: this is again an issue due to the limited size of the NMOS. While in standard packaged level requirements for 2 kV HBM and 200 V MM the 750 V CDM protection level can be achieved with a reasonable clamp size of ~1,00,000 μm^2, the system-level requirements for power-on and ESD current level tolerance ten or more times higher typically make active clamp design impractical.
(iv) High-speed limitations: in the case of fast transient pins, the clamp cannot support normal operation conditions and will drain excessive current through the protected pin.
(v) ESD PLUS and ESD MINUS bus: organizing an ESD network with an active clamp requires a low-resistive ESD bus structure.
(vi) Cross-talk: in the ESD protection network, pins with shared active slave clamps will detect some signal due to parasitic coupling in diodes and the finite resistance of the metal interconnects. Thus, this clamp solution is not applicable to the multiple channel low-noise amplifiers with multiple inputs and outputs, for example.

Thus, in spite of the elegance of this solution and the provided low clamping voltage in analog and high-speed designs, high-voltage and multiple domain designs utilize snapback clamps that employ conductivity modulation mechanisms (discussed in Chapter 2) and are based on ESD devices (discussed in Chapter 3). These snapback clamps are discussed in the following sections of this chapter.

Other types of active clamp solutions based on NPN BJT and PMOS devices, as well the ESD protection network with active clamps, are discussed in the first section of Chapter 5.

4.2 Low-Voltage Clamps with Internal Blocking Junction Reference or d*V*/d*t* Turn-on

4.2.1 Snapback NMOS Clamps

4.2.1.1 Ground-Referenced Snapback NMOS

Snapback NMOS is one of the most popular solutions for local ESD protection that can be implemented in any CMOS process as a "free" ESD device. The alternative name for this clamp often used in publications is grounded gate NMOS (ggNMOS).

Depending on the technology used, the snapback NMOS device is usually capable of supporting up to a 2–5 mA/μm current density in high-current mode. Respectively, the size of the snapback NMOS ESD device in the clamp usually varies from 200 to 600 μm. At the pad dimension of ~60–100 μm, the specified clamp width with all guard rings is expected to be within ~100 μm. Therefore, the layout is implemented as an array with multiple fingers of 40–70 μm width each.

An example of the 5 V snapback NMOS clamp schematic and layout view for a 0.5 μm process is presented in Fig. 4.2. The circuit for the clamp includes a gate resistor and an optional reverse path diode.

In principle, operation of the clamp is based upon parasitic NPN avalanche–injection (Chapter 2) controlled by the channel current incoming into the multiplication region (Chapter 3). Therefore, there are three controlling factors for clamp turn-on:

(i) drain avalanche breakdown
(ii) dynamic coupling of the gate
(iii) displacement current in the parasitic n–p–n structure

4.2.1.2 Gate Coupling

A gate resistor of a sufficiently high value can be selected to enable turn-on of the clamp due to the gate capacitive coupling effect by the fast voltage rise on the protected pad. This d*V*/d*t* effect is the result of the triggering voltage dependence on the level of channel current.

In addition to the gate coupling effect the lengthy drain ballasting region already results in a, respectively, higher internal drain-bulk capacitance, in comparison with the bulk-source capacitance, the internal divider creates a displacement substrate current. Since the critical voltage for current instability in a silicon NMOS device usually depends upon both the channel current and substrate potential (Chapter 3), the triggering voltage in transient conditions is usually achieved at significantly lower level than the dc breakdown voltage. For example, for the clamp in Fig. 4.2, the dc breakdown voltage is ~13 V.

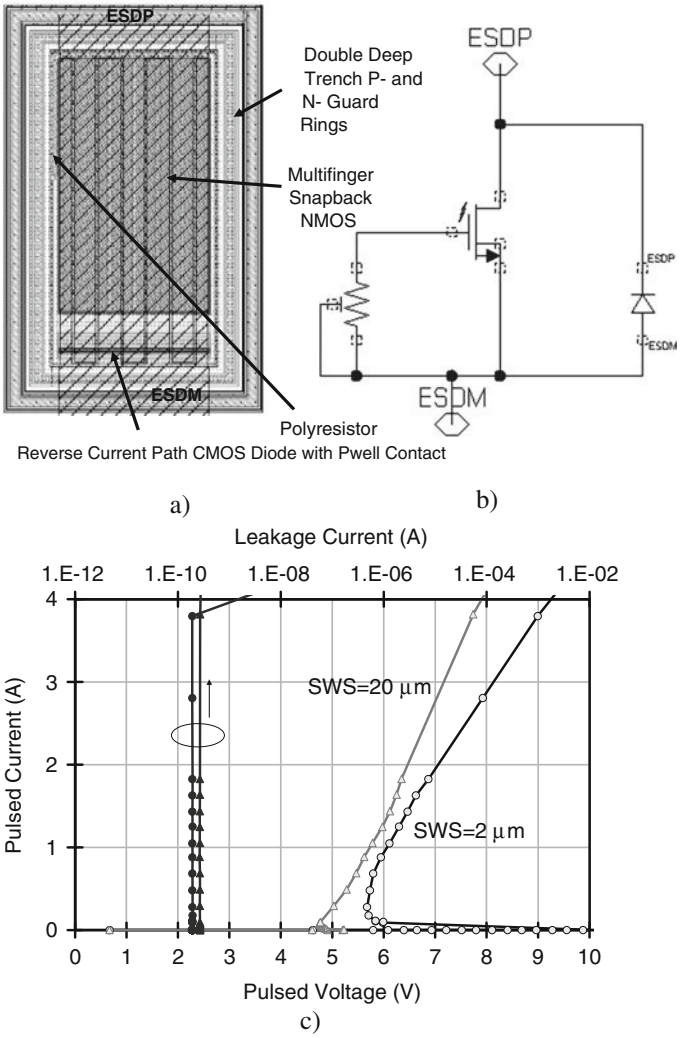


Fig. 4.2 Layout view (**a**) and circuit diagram (**b**) of the snapback NMOS clamp and measured TLP characteristics (**c**) for two different well-to-source parameters SWS (**c**)

4.2.1.3 Displacement Current Effect

The second factor controlling the clamp turn-on is the intrinsic base resistance of the parasitic n–p–n device. In the clamp, the source and well tap p+-region are connected at the metal. Therefore, the internal base resistance of the parasitic n–p–n structure is the critical parameter for the avalanche–injection instability boundary of the final distributed array. In principle, in the case of R_G=0, an additional component of the dV/dt effect can be observed as a decrease in the triggering voltage

with the decrease of the pulse rise time. This effect could be the result of a pure displacement current effect in the parasitic n–p–n structure. However, in the case of long finger NMOS devices formed in non-silicided processes or a silicide blocked gate region, the gate coupling can still be a dominant factor. In these conditions, the distributed resistance of the gate allows significant gate potential increase in transient conditions, even if the polygate has a metal connection to the source at the end of each finger.

Exclusion of the silicide region from the gate is usually a side effect of the implementation of a silicide blocking region. One of the practical implementation issues is the formation of a narrow silicide stripe region along the gate edge. This is possible if the silicide blocking mask is aligned with the side edge of the polygate drain. The most practical measure to avoid such a silicide sliver formation is to align the silicide blocking mask to the middle of the gate.

Since parasitic base resistance in the n–p–n structure has a significant impact on the instability boundary, there are many practical ways to make clamp layout design advantageous. The clamp is usually designed as an array of non-butted NMOS fingers with spaced well tap contact regions to accumulate the necessary total width (Fig. 4.2a).

There are several ways to implement the Pwell connection to increase the base resistance of the internal NPN. In the most simplistic cases, a p+ ring that is already required for latch-up isolation can be drawn at some distance from the NMOS source-drain regions. In this case, the middle part of the array will have a significantly larger spacing, in comparison with the corners of the array, with a corresponding increase in base resistance. The increase of the base resistance will reduce the triggering voltage (Fig. 3.2c) in accordance with avalanche–injection conductivity modulation in n–p–n structures (Chapter 3).

In spite of the concern of possible current crowding, the total finger width may still be utilized in high-current operation of the structure. Additional current ballasting can be achieved using comb ballasting region design [72].

It can be shown that in such design, turn on of the middle of the device in the conductivity modulation state will cause the remaining part of the device to engage. Thus, the device will work as a single entity. In the case of isolated CMOS processes, BiCMOS processes, or SOI processes, either a local or soft-well connection can be realized.

Another important note can be made for dV/dt-triggered clamps. The transient gate-coupling effect might become a problem in high-speed I/O circuits. At a data rate of $\sim$10 Mb/s, the rise time of the signal at the pad is already faster than in standard HBM and MM ESD pulses. In spite of the high slew rate, the voltage amplitude cannot trigger the ESD clamp into snapback mode but the coupling of the gate may result in significant current through the clamp, thus deteriorating the high-speed performance of the I/O pin. Compact modeling with extracted parasitic resistance and capacitance from metallization is helpful but cannot guarantee the end result due unaccounted for parasitic n–p–n structures in typical NMOS models, for example, in the widely used BSIM3v3.

Alternatively the reduced triggering voltage can be achieved by decrease of the breakdown voltage with additional ESD implant [73].

4.2.1.4 Reverse Path Protection

For every local clamp, the most typical requirement is dual-direction protection. Usually, the voltage specification at negative bias is low and therefore a simple diode can provide the desired result. An advantage of the snapback NMOS clamp is the option to integrate the diode inside the clamp. The diode can be easily formed by adding an n^+-diffusion region near the p^+-diffusion in Pwell. The only important consideration is to ensure appropriate metallization for such a diode.

The reverse path diode is a typical internal clamp component. In principle, the diode is already presented in the clamp as a body diode and is formed by the p-tap p+ region and the n+-drain diffusion region. However, due to the large distance between these two regions, the diode resistance is rather high. Thus, an additional n+-diffusion region is usually connected to the pad inside the cell in the proximity of the p+-diffusion to form the diode. In the clamp (Fig. 4.2a), this region is placed below the NMOS fingers. In this case, appropriate metallization should be provided for the high-current level.

Respectively, there is the option to separate the p^+-tap region from the drain-source area by the n+ reverse path diode region. This brings a new topological degree of freedom for control of the base resistance of the parasitic n–p–n structure on the device level. The degree of impact for such a design approach depends on the Pwell profile. The diode n^+-region provides additional depletion in the Pwell, thus controlling the equivalent resistance to a higher degree.

Often, the reverse path diode can be implemented as separate fully isolated structure. Depending on the process, a 50–100 μm width is usually sufficient for a standard minimum package level ESD spec (2 kV HBM, 200 V MM).

4.2.1.5 Isolated Snapback NMOS

The snapback NMOS clamp is usually designed to target the ESD protection window using the transient triggering effect. In this case, the triggering voltage is realized below the expected triggering voltage for internal circuit devices with low gate potential during ESD stress. However, if the coupling of the internal device is high, protection of, for example, an open drain circuit creates a significant challenge. In this case, a device of the minimum gate length may provide the triggering voltage at a rather low level, thus taking over the ESD current path. Usually, CMOS process provides no degree of freedom in changing the blocking drain junction in the snapback NMOS.

In the case of isolated CMOS process, the isolated snapback NMOS clamp can be designed. For substrate isolation, a deep N-region (NBL, NISO, DeepNwell, Chapter 3) is used in combination with appropriate spacing for lateral isolation. This clamp is required for non-ground-referenced pins or high-side pins, for example, BOOST PIN in DC–DC converters (Chapter 6).

The main design principles of isolated snapback NMOS clamps are the same as those of regular snapback NMOS clamps. However, in the case of isolated devices, an additional advantage can be derived from the reduced well resistance (Fig. 4.3). The middle part of the clamp is the same as in non-isolated snapback clamp layout design (Fig. 4.2a). An additional isolation N-ring and an additional N-epi region connection are required.

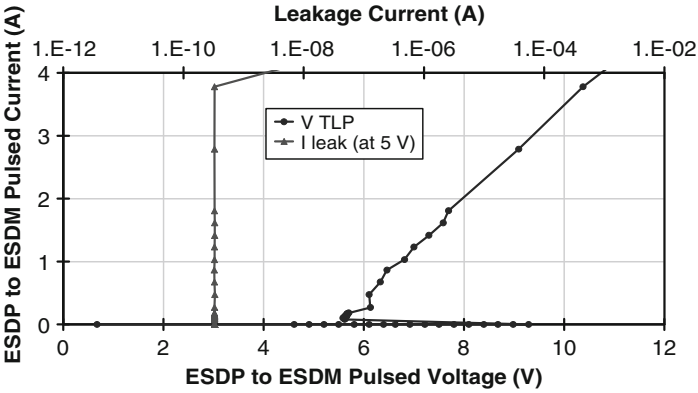


Fig. 4.3 Measured TLP characteristics of the isolated snapback NMOS clamp

In BiCMOS processes, the buried layers can be used to control characteristics of the NMOS-based clamps.

As a result, sheet resistance of the Pwell isolation can be higher if either the P-buried layer is not used or the N-buried layer is used instead. In this case, the triggering voltage of the clamp can be achieved at a much lower level. This can be illustrated by a comparison of the TLP I–V characteristics for the non-isolated clamp (Fig. 4.2c) and its isolated version (Fig. 4.3).

Both snapback NMOS and isolated snapback NMOS rely on the dV/dt effect in order to bring the triggering voltage of the clamp inside the ESD protection voltage range.

In addition to the already mentioned drawbacks of the transient triggered clamps, the dV/dt effect also has a strong dependence upon the load realized by the internal circuit (Chapter 5). The lowering of the rise time for the ESD pulse is simple to understand, taking into account a possible additional capacitance of the internal circuit which is effectively connected in parallel to the ESD protection structure. This capacitance can be significant compared to the pulse forming capacitance, in which case, the rise time of ESD pulse can be significantly increased.

The alternative method to adjust the triggering voltage involves a voltage reference component in the drain-gate circuit or alteration of the NMOS device toward lower breakdown voltages.

In the case of the typical CMOS process, options for these changes are quite limited and very process dependent. It appears the only practical self-aligned approach

is to use longer gate length in the I/O components and minimum gate length in the ESD NMOS clamp. Other possibilities include partly blocking the nldd implants from core low-voltage devices in the case of the DGO CMOS process.

In the BiCMOS process, the PBODY mask and other shallow implants that support the bipolar part of the process may be used to reduce the breakdown voltage to bring it inside the ESD protection window.

All these measures require experimental validation to confirm compatibility with the given process. This statement is true for any snapback device solution.

Some I/O libraries are designed using snapback NMOS for the dual purpose as both an I/O buffer component and an ESD protection clamp element, thus introducing self-protection capability.

Various styles of NMOS clamp design can be implemented based on designer preference or the process specifics. For example, based upon comparison of the experimental results in Figs. 4.2 and 4.3, no difference is found between vertical or lateral finger directions in the clamp, as long as a number of backend design rules are followed. Another typical use of the isolated NMOS clamp is in the stacked clamp for higher-voltage tolerance [74, 75].

4.2.1.6 The 40× Rule for Backend

For the ESD condition, the "40×" rule for metallization is usually followed. The rule simply reflects a correlation factor between the critical density for metallization burnout in ESD pulse conditions and the experimentally measured electromigration rules of the process. The exact correlation factor can be experimentally determined for the given process and, perhaps, may not be exactly 40. The rule simply suggests that the safe current density level through the contact, via, and metal layer region can be estimated by multiplying the electromigration current density maximum rating by the number 40.

According to design experience, this rule is validated for at least 0.18–0.5 μm process technologies, it may require additional verification for 90–32 nm process technologies with new materials [76].

In an ideal design, clamp metallization is expected to provide the same metal resistance to each point of the device. This is usually achieved by using comb-like metallization, applying an approach similar to the metallization of power arrays. Power arrays will be discussed later in this chapter.

One of the assumptions of snapback NMOS cell design is that device turn-on may happen in part of the cell, but eventually all the fingers are switched into conductivity modulation mode. In other words, a turn-on domino effect is realized inside the clamp. If a local part of the clamp or a part of a finger turns-on into the high-conductivity state, the lateral diffusion of the carriers provides significant initial current density in the avalanche multiplication region. Thus, spreading of the conductivity modulation process occurs along the whole finger width and the ESD cell. However, if each clamp finger is fully isolated, then there is no coupling between the fingers. In this case, the first finger that turns-on takes over all the ESD current, thus blocking turn-on of the other fingers. This effect may produce multiple

S-shaped regions that correspond to the contribution of the active fingers [30]. This effect depends on the ratio between the triggering voltage V_{T1} and the critical voltage for maximum current V_{T2}. If V_{T2} is exceeds V_{T1}, then the additional fingers may turn on after current saturation. However, it should be taken into account that the secondary turn-on may occur in conditions of much higher rise time. Therefore, at the same equal conditions, the triggering voltage V_{T1} for remaining fingers can be substantially higher. Nevertheless, if $V_{T2} < V_{T1}$, then the burnout of the first finger can be expected.

A proper design of both backend metallization routing and well tap connection and gate coupling are important to avoid the *multifinger turn-on* effect. If the process does not allow any alternatives, additional backend ballasting has been suggested in [30].

An additional design parameter is contact enclosure in the diffusion region. One important thing to consider is that process design rules are usually engineered to support the current density for normal device operation. Taking into account process variability and lateral diffusion, it is logical to expect that the minimum contact enclosure might not be optimal for a high-current density ESD device. If at the ESD current level the contact region is not quasineutral, then an excessive power generation can cause an elevated local temperature at the contact which may result in irreversible processes in the metallurgical contact structure. One of the adjacent aspects of pertaining design solutions is implementation of the appropriate contact enclosure in the lateral regions of the finger.

4.2.2 Transient-Triggered PMOS Clamp

Similar design concepts can be applied to the PMOS-based clamp [77]. Unlike in the snapback NMOS clamp, the parasitic p–n–p structure in a low-voltage PMOS clamp provides either absence of or moderate S-shaped *I–V* characteristics. This phenomenon is a direct consequence of the different level of space charge compensation in the p–n–p structure versus the n–p–n structure determined by the physical difference in the carrier mobility of the silicon semiconductor material.

Respectively, the higher holding voltage is realized due to the ratio of nobilities of the majority and minority carriers (Chapter 2) under the low gain of the parasitic PNP structure.

However, the basic avalanche–injection process in PMOS is the same as that in the NMOS, so the negative differential resistance effect is a strong function of the implant profiles of the device. Usually, negative differential resistance can be observed at a low current level in short channel PMOS devices with a gate length below 0.18 μm, thus revealing the general nature of the effect.

At the same time, absence of negative differential resistance provides an advantage. In the case of the PMOS clamp, current stratification is absent, eliminating the need for the drain ballasting region. As a result, the clamp can be composed within the same area, but with a much higher total width of the array.

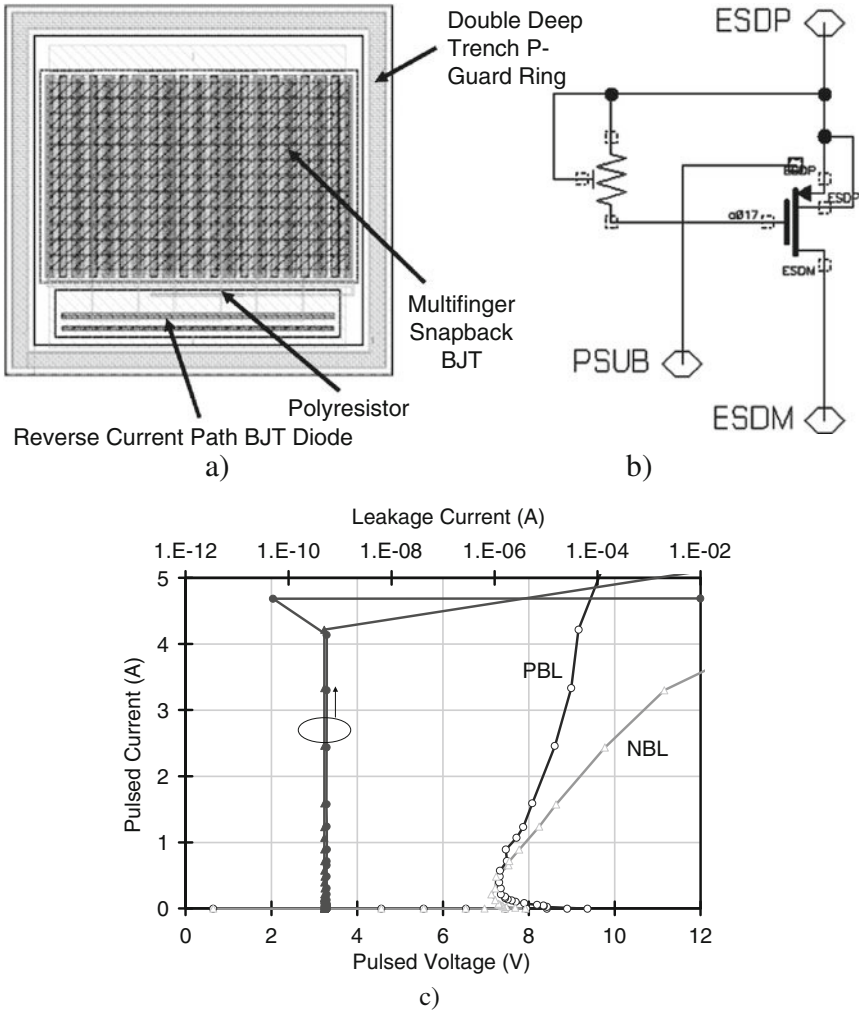


Fig. 4.4 Topology of the 5 V PMOS clamp (**a**), circuit diagram (**b**), and measured TLP characteristics (**c**) for the clamp based upon a PMOS device with an NBL layer and PBL layer

The design and TLP characteristics of a W=1 mm total width 5 V PMOS clamp are presented in Fig. 4.4. This clamp is implemented in the same BCD process as the previously discussed NMOS clamps. The ring around the source-drain PMOS arrays forms the N-well contact region.

With the same footprint as snapback NMOS, this clamp provides similar high-current operation. The turn-on point is controlled by the high Nwell resistance path and the gate-coupling resistor R_G is connected to the pad.

The potential use of the clamp depends on the ratio between breakdown voltages which can be achieved in NMOS and PMOS structures. In terms of transient effects, the clamp has the same peculiarities as the transient triggered snapback NMOS clamp.

In principle for the CMOS the clamp characteristics are dependent upon substrate type due to internal base resistance impact. The impact of the substrate resistivity is studied in [76].

4.2.3 10 V FOX Snapback Device

The next simple self-aligned ESD clamp for the CMOS part of the process is based upon the field oxide device (Chapter 3). In principle, this clamp eliminates the gate regions and subsequent gate-coupling effects, leaving only the free lateral NPN-BJT structure.

The clamp can reuse the same topology and backend metallization as the previously designed snapback NMOS clamps. Thus, a simple transformation rule replaces the former polygate region with the minimum dimension thick field oxide isolation region that can also be applied on the cell level. An example of the ESD clamp layout and schematic view is presented in Fig. 4.5.

In principle, the device in the clamp presents an implementation of a free lateral NPN by reusing CMOS diffusion regions. This NPN has a long P-base and thus a rather low current gain before breakdown. The high internal base resistance is controlled by the Pwell-source and active region spacing (usually, naming conventions for device contacts reuse the names of corresponding regions in the ancestor snapback NMOS).

The major change, in comparison with the snapback NMOS (Fig. 4.2c), is the higher triggering and holding voltage characteristics (Fig. 4.5c).

The major advantage of this clamp is based upon the device structure itself. The device eliminates limitations on the voltage tolerance related to gate dielectric breakdown and hot carrier degradation, combined with a predictable self-aligned blocking junction formed by the n+-diffusion and Pwell.

Thus, this clamp is an efficient solution for a case where voltage tolerance of the analog circuit pin exceeds the CMOS voltage level and thus protection cannot be achieved by the snapback NMOS cell.

Most typical cases are the EPPROM “write” and “erase” pins or stacked I/O protection.

The same topological control of the internal base resistance can be implemented using local, partly isolated, or n+-diffusion blocked options for Pwell tap placement acting as a base for the parasitic n–p–n structure.

For the CMOS process, device turn-on will be based on the given n+-Pwell breakdown voltage. Presence of the poly or metal field electrode on top of the isolation region can be used for turn-on voltage control. However, such a solution is usually rather marginal due to process design rules. Thus, the basic method in targeting the

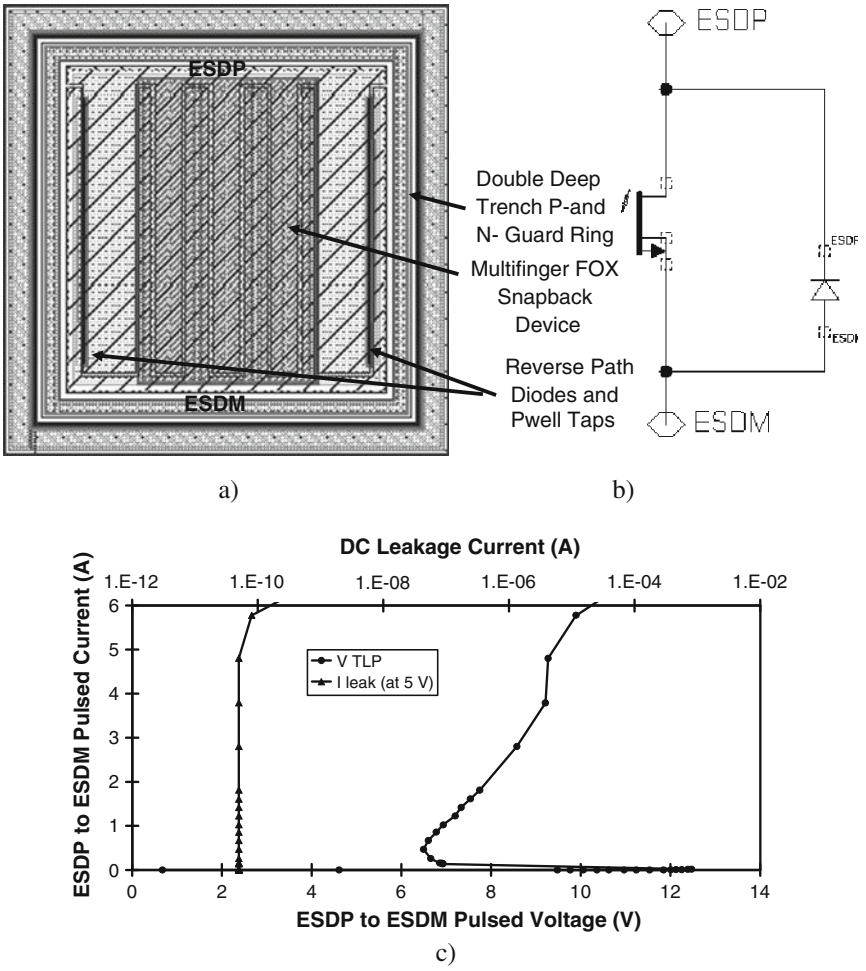


Fig. 4.5 Topology of the FOX clamp (**a**), circuit diagram (**b**), and measured TLP characteristics (**c**)

ESD protection window is to utilize the d*V*/d*t* effect due to the displacement current. An alternative method is to alter the base–collector (Pwell-source) blocking junction (Chapter 3).

The base (Pwell) resistor to the ground enables additional lowering of the turn-on voltage. In the case of non-isolated CMOS processes, the Pwell is softly connected to the p-substrate and thus shunts the resistor. In this case, the substrate pump method is realized [78]. In the isolated clamp version, the well diffusion connected to the source through the resistor acts similarly to the external base resistor in the NPN clamp in reducing the triggering voltage (Chapter 2). The Pwell (base) contact can also be used to inject the voltage-referenced base current.

4.2.4 LVTSCR and FOX-SCR Clamps

Similarly to snapback NMOS and FOX devices, the LVTSCR and FOX-SCR clamps can be implemented using the same footprint, topology, and design features to adjust the triggering characteristics and provide an appropriate level of width scaling. Transformation rules for the device can be applied on the cell topology level, simply replacing the internal snapback NMOS device array with corresponding SCR devices, as described in Chapter 3.

An example of the LVTSCR clamp is presented in Fig. 4.6. Since the SCR device usually provides an on-state current approximately one order of magnitude higher, the clamp size can be significantly reduced (ultimately down to a 25–40 μm-width ESD device).

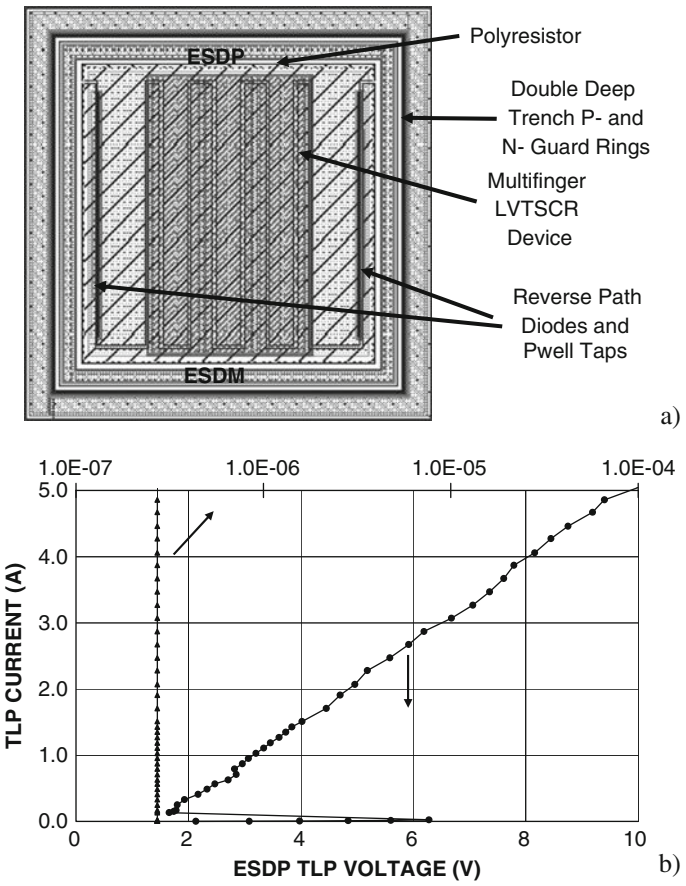


Fig. 4.6 Topology of the LVTSCR clamp (**a**) and measured TLP characteristics (**b**)

In the case of the LVTSCR clamp [51], the gate coupling is used to trigger the LVTSCR into high-conductivity state similarly to triggering in snapback NMOS. In field oxide SCR, the advantage of displacement current in the parasitic NPN and PNP devices is used to turn the clamp into high-conductivity mode.

In spite of the advantage in high-current conduction, the disadvantage of the SCR clamps is the rather low holding voltage. This is already discussed in Chapter 3 on the device level. There are several solutions for implementation of the higher holding voltage levels using both device and subcircuit components.

In general, the low holding voltage of the clamp creates a major problem for pins sensitive to transient latch-up. To eliminate this problem, several solutions have been suggested that involve implementation of certain high-current positive feedback on the clamp in order to increase the holding voltage above the power supply level. This is discussed in Section 4.2.5.

According to the original design concept of the device, the triggering of the clamp is first based upon the snapback NMOS turn-on. Then, when a certain current density level is achieved during turn-on, the p+-emitter region becomes positively biased over 0.8 V, and direct injection from the p+-emitter region initiates double injection conductivity modulation (Chapter 2).

However, in principle, the clamp can be designed to fulfill the double S-shaped characteristics. The first snapback region corresponds to the NMOS avalanche–injection operation. Then, the condition for SCR double injection conductivity modulation is achieved only at a rather high snapback current level. This case is usually encountered at excessive p+-emitter isolation.

Alternatively, with p-emitter isolation reduction, the parasitic p–n–p structure can become dominant and the device may turn-on directly into the double injection mode. Repositioning the p+-emitter closer to the gate region or disconnecting the drain can achieve this. In the last case, the device will be turned on by the punch through effect even if it skips the avalanche breakdown stage.

Another typical challenge, especially for FOX-SCR devices, is the triggering voltage control. Section 4.2.5 presents some practical methods of triggering characteristics control using a special control circuit connected to the clamp.

4.2.5 High Holding Voltage LVTSCR Clamps

Two methods of increasing the holding voltage have been suggested: circuit and topology-based device solutions.

4.2.5.1 High Holding Voltage Cell Topology

In the LVTSCR device, positive feedback in the injection current from the emitter and the source junction results in the formation of a quasi-neutral region, as a result of mutual space charge neutralization of the injected electrons and holes. The final

holding voltage after triggering is an integral function of the resultant electric field distribution in the discharge region.

The method of increasing the minimum holding voltage as a result of the change in carrier balance on the device level has been proposed in [37, 79]. Carrier balance control has been achieved by implementation of the different levels of emitter junction isolation on the device topology level. To achieve the controllable imbalance in the injected charge, the p-emitter length is reduced down to a minimum feature dimension while maintaining proper device spacing. The hole injection current is limited, thus defining the holding voltage.

Similar effect has been reported in [80]. The effect was achieved by using *N*-isolation region to change the carrier balance in the device.

A similar effect can be achieved on the clamp layout level by changing the emitter isolation topology. One of the examples includes implementing the p+-emitter by interdigitated islands surrounded as the n+-diffusion and increasing the distance between the p+-emitter and the source, or reversing the positions of the drain and the p+-emitter.

In an extreme case, if the p+-emitter is over-isolated, the snapback NMOS operation current density level may not produce a p+-emitter voltage drop sufficient to overcome the potential barrier of ~0.6–0.8 V. Thus, no SCR-specific double-injection effect will be produced in the device. Such an SCR device will be physically equivalent to a snapback NMOS with an additional Nwell saturation resistor in the wide current range.

This topological method has been validated by numerical simulation and followed by pulsed measurements in [81]. The original device structures of the LVTSCR were created using a calibrated 0.18 μm CMOS TCAD process simulation flow. The 2D structures were generated using process [82] and device simulators [83]. Cross-sections for the LVTSCR and NMOS structures used in the simulation analysis are presented in Fig. 4.7a, b. The emitter isolation effect is presented in Fig. 4.7c. A comparison between snapback NMOS and SCR with a reduced emitter region demonstrates a higher current level from the SCR device optimized for 3.3 V-tolerant ESD protection.

The measured data confirmed simulation results. A strong dependence was observed of the holding voltage on the structure and the length of the emitter region (parameter *LN*, Fig. 4.7a). The on-state current is dependent upon the length of the emitter. As the emitter region is reduced, the holding voltage increases and the saturation current decreases. Stable triggering characteristics (at ~10 V), with an intermediate value for the holding voltage, are obtained in a 0.5–3 μm range of the length of the emitter region. The holding voltage range achieved is between the conventional LVTSCR holding voltage value (~1.5–2 V) and the snapback NMOS holding voltage value (~5–6 V) (Fig. 4.8). A comparison of the measured snapback characteristics for the grounded gate snapback NMOS clamp and the LVTSCR clamp with an emitter region that delivered similar holding voltages is presented in Fig. 4.8b.

Thus, the high holding voltage in the LVTSCR provides both higher current in the saturation region after triggering and, at the same time, higher critical power

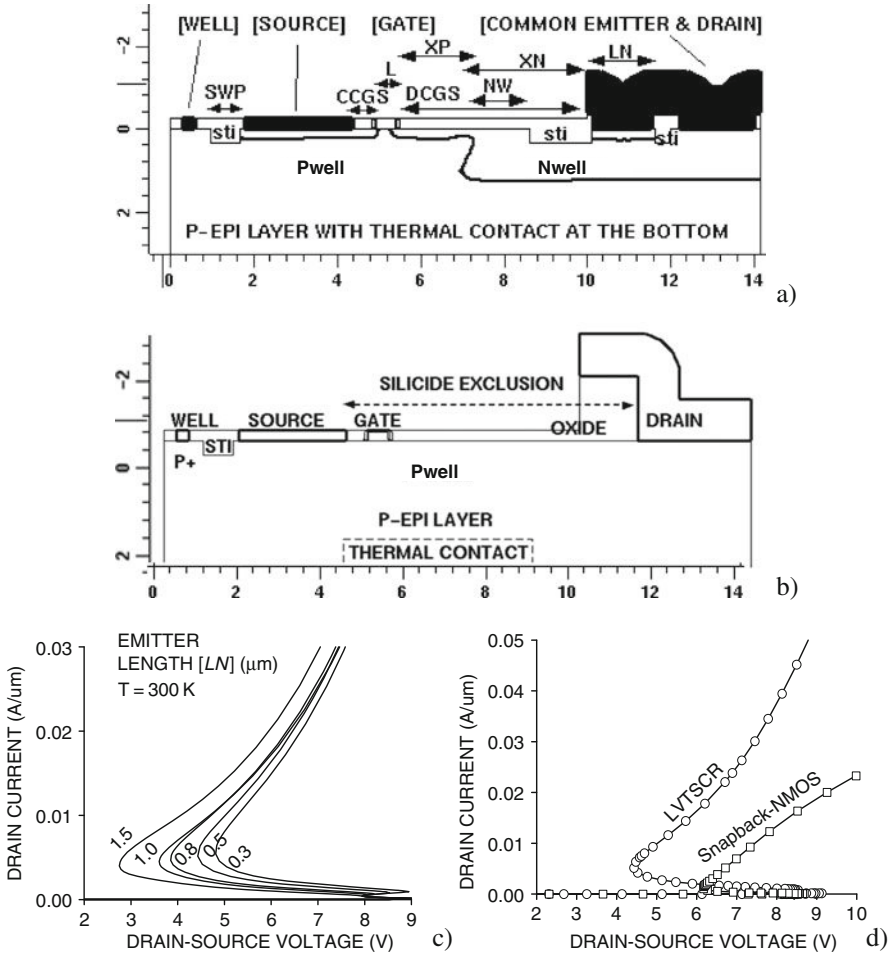


Fig. 4.7 Simulation cross-section for LVTSCR (**a**) and NMOS structure (**b**). Isothermal *I–V* characteristics for different emitter area lengths (**c**) and comparison of the LVTSCR structure with a 0.5 μm emitter region L_N and the snapback NMOS structure (**d**)

(that corresponds to the final point on the TLP curve in Fig. 4.8b). Similar results have been demonstrated for FOX-SCR clamps.

4.2.5.2 Clamp-Level High Holding Voltage Using P-Emitter De-biasing

An alternative and a more elegant way to solve the holding voltage problem on the clamp design level in SCR structures has been proposed in [81]. The emitter injection control approach has been suggested and implemented using a new structure (Fig. 4.9). This approach involves the high-current de-biasing diode structures.

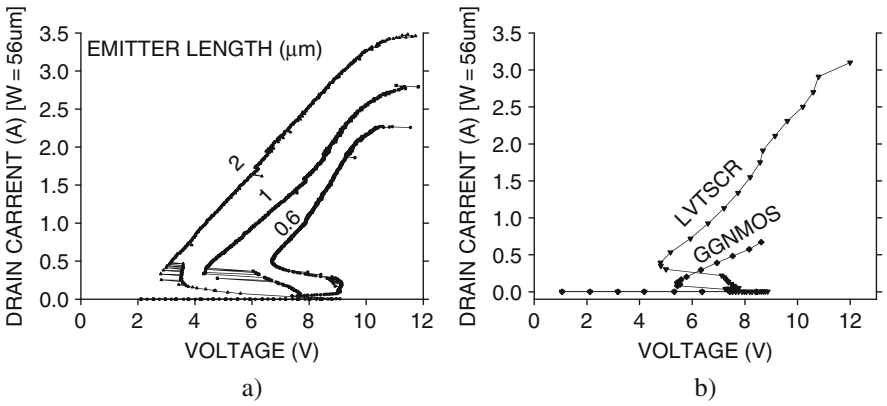


Fig. 4.8 Measured TLP *I–V* characteristics of LVTSCR: (**a**) for different emitter region lengths; (**b**) comparison of NMOS and LVTSCR with the same holding voltage; the last point of each curve corresponds to the beginning of soft-leakage degradation

The diode structures can be implemented as stand-alone or as internal components inside an N-well. They create an internal circuit that provides an additional voltage drop in the p+-emitter circuit (Fig. 4.9a, b).

Due to the additional forward biased junctions, the emitter injection is limited after triggering by the voltage drop on the diodes, thus cutting off emitter injection at the corresponding voltage level (Fig. 4.9c).

In a conventional LVTSCR, after the ESD triggering, the balance between injection from the emitter and the source junctions creates a quasi-neutral electron–hole plasma region that produces a low-voltage drop in the on state. The charge balance of the injected carriers determines the voltage drop. Therefore, by changing this balance in high injection conditions, an impact is expected on the holding voltage.

To implement this physical principle, the LVTSCR emitter has been designed with a separated contact. This emitter contact was connected not directly to the protected pad, but to a string of forward-biased diodes instead. In this case, the holding voltage decrease in SCR mode is limited by closing the emitter junction below the reference voltage drop on the emitter diodes. Consequently, this effect results in a shift in the balance between injected carriers and a change in the level of the carrier space charge neutralization, increase of the electric field and, finally, creation of a higher source-drain voltage drop.

Thus, an intermediate, mixed regime can be obtained with a holding voltage range between double injection and avalanche–injection. In general, selecting the proper number of diode structures in the emitter circuit can control the holding voltage.

In [81], has been shown that even at the same holding voltage as that of snapback NMOS, the high holding voltage LVSCR is still several times more efficient due to its low on-state resistance (Fig. 4.9c). This makes the diode component a

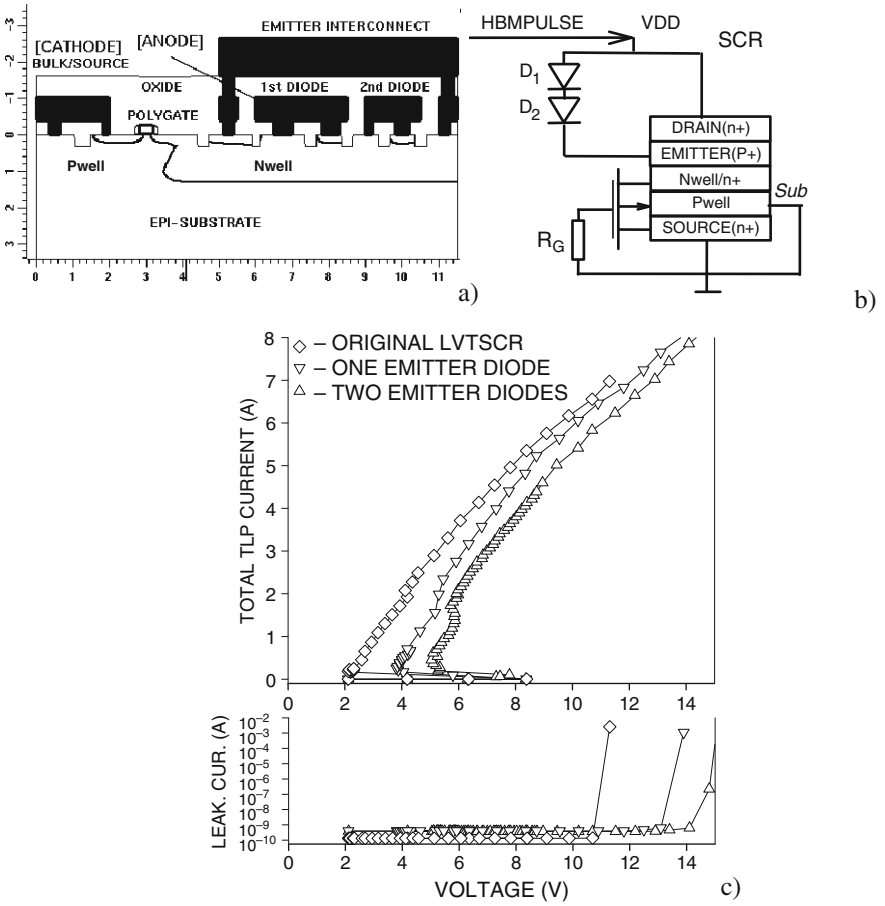


Fig. 4.9 Cross-section of the LVTSCR drawn using a physical process simulation (metallurgical junctions are shown) (**a**), its simplified equivalent circuit (**b**), and calculated isothermal triggering *I–V* characteristic (**c**). Experimental TLP data for different LVTSCR design with 100 μm total contact width, presenting the reference LVTSCR structure with a common emitter/drain contact, LVTSCR with one and two diodes in the emitter line inside the N-well (**c**)

very valuable protection solution for different system applications, including charge cable specified interface applications.

4.2.6 Triggering Characteristics Control in SCR Clamps

In addition to holding voltage control, another important clamp parameter is the triggering voltage. This section presents the most advanced method to provide such control, based upon the mixed device-circuit concept. In this section, the approach is applied to a generic SCR p–n–p–n clamp, which is based on devices similar

to corresponding structures discussed in Chapter 2. The design of such a device assumes the presence of two emitters, p+ and n+, isolated by corresponding n-base and p-base regions. Both bases form the blocking p–n junction that provides the high-voltage tolerance of the device. The bases have p+ and n+ contact regions with corresponding control electrodes.

4.2.6.1 Mixed Device-Circuit Concept

The mixed device-circuit ESD solution has been first proposed in [85]. Similar principle has been used in recent study [84]. It is based upon a different principle than conventional ESD clamps. Instead of solving the problem of targeting the ESD protection window in both the voltage and rise-time domains, the mixed device-circuit approach provides for different triggering characteristics of the device under ESD and normal operating conditions.

The snapback ESD device is first modified to provide access to a terminal that the active circuit will control. The snapback device is required to have (i) a very low triggering voltage in the case of a positive bias at, or forced current through, the control electrode, and (ii) a triggering voltage that is higher than any expected voltage in the case of a grounded control electrode. Requirements for the active control circuit are as follows: (i) to provide control electrode biasing or current forcing during the first part of an ESD event (∼20 ns), in order to turn-on the ESD device; (ii) to ground the control electrode after the ESD device is operating in snapback mode (after ∼100 ns); (iii) always turn-off (ground) the control electrode if VDD (∼5 V) is generated.

4.2.6.2 Practical Implementation of the Concept for the Case of 130 nm Process

If voltage tolerance above the limitations of the NMOS gate oxide is required, an SCR can be used, thus excluding the gate oxide region. However, in this case, the triggering voltage is rather high, especially in CMOS processes of 0.13 μm and below.

With the scaling of CMOS processes down to 60–45 nm the voltage tolerance becomes rather low. One of the logical solutions is to use stacked forward-biased diodes to protect the I/O pins. However, the high temperature dependence and high-frequency I/O pin requirements for low-capacitance ESD protection circuits demand a more advanced approach.

For applications in which the dual-diode protection circuit cannot be used, SCR-based solutions [86] provide the highest level of protection per unit capacitance.

Since the blocking junctions in the SCR device have a rather high breakdown voltage of ∼5–7 V, trigger circuits are needed to turn on SCR's quickly and at sufficiently low voltages.

The diode-triggered SCR (DTSCR) [87] (Fig. 4.10a) and grounded gate NMOS-triggered SCR (GGSCR) [38] (Fig. 4.10b) are suitable for protecting low-voltage I/Os. The DTSCR is favored over the grounded gate SCR GGSCR because its

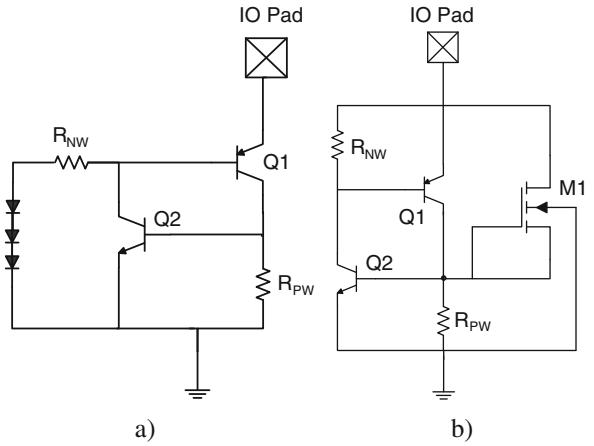


Fig. 4.10 Circuit schematic of the DTSCR (**a**) and GGSCR (**b**)

trigger voltage, V_{T1}, is easily adjusted by altering the number of diodes in the trigger circuit. Additionally, in advanced technology nodes, special process steps, e.g., silicide-block, may be required for the GGNMOS embedded in the GGSCR. The downside of the DTSCR is its significantly higher leakage current.

A different approach has been proposed in [88], featuring the dual-base triggered SCR. The proposed solution combined the best features of the GGSCR [38] and the DTSCR [87]; it has low off-state leakage current and an easily adjustable trigger voltage. The turn-on time was virtually identical to that of an optimized DTSCR designed in the same technology.

A major step in the philosophy of new design was to use the informative VDD signal produced by the internal circuit, this approach was proposed in [85] and elaborated further in [89, 90]. Although this solution hierarchically partly belongs to the ESD protection network level presented in Chapter 5, it will be discussed in this chapter to allow comparative analysis.

The core VDD is used as a control signal to prevent triggering and reduce leakage current while the chip is powered, this is a successful application of the high-voltage ESD protection concept introduced in [85] to a low-voltage ESD protection scenario.

The trigger circuit permits one to design for low V_{T1} values during ESD without compromising resilience against false triggering under normal operating conditions, because the circuit's V_{T1} automatically increases when the chip is powered up.

4.2.6.3 Principle of Dual-Base Control Operation

The dual-base triggered SCR clamp is shown in Fig. 4.11. It consists of an 11-transistor trigger circuit and an SCR [91]. The trigger circuit is composed of a control circuit (M0–M3), a voltage divider (M4 and *R*0), and a chain of three

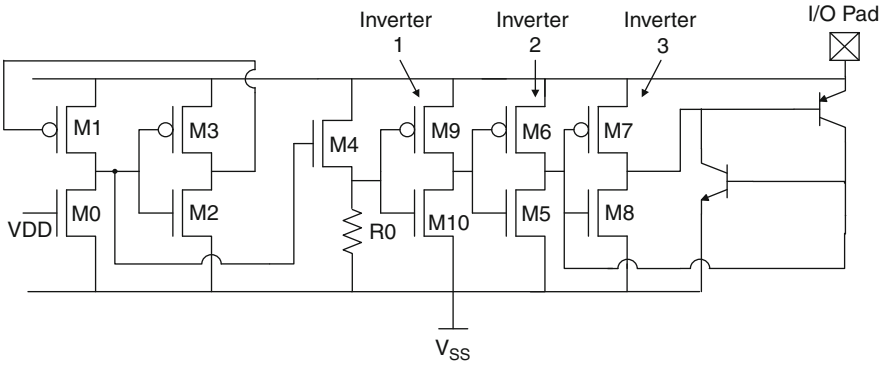


Fig. 4.11 Circuit schematic for dual-base triggered SCR clamp

inverters (M5–M10). The trigger circuit consumes a relatively small layout area as all of its transistors are nearly of minimum size except for M6 and M8. Thick-oxide I/O transistors are used throughout the circuit to reduce leakage current and increase the ESD robustness of the trigger circuit.

The Pwell of the SCR is connected to the output of the second inverter and its Nwell is connected to the output of the third inverter. These connections and the resulting dual-base triggering action are similar to that of the RC-triggered SCR [92]. However, the triggering voltage of the RC-triggered SCR is not adjustable, unless significant design changes are to be made to the trigger circuit.

Furthermore, the dual-base triggered SCR uses VDD as a control signal to prevent triggering and reduce leakage current while the chip is powered. In principle, any power supply bus can provide the control signal; however, application-specific considerations (e.g., power-on sequence, bus routing) dictate whether core VDD or I/O VDD should be used.

While the chip is powered, VDD is high and M0 pulls the gate of M4 down to ground. Consequently, the output of the first inverter is high, the output of the second inverter is low, and the output of the third inverter is high. The SCR N-well is pulled up to the I/O pad voltage and the P-well is pulled down to ground, thereby shorting out the base–emitter junctions of the PNP and NPN transistors. This keeps the SCR in the off state and minimizes its leakage current.

During ESD conditions, VDD may be closely coupled to VSS, depending on the topology of the ESD protection network and the amount of decoupling capacitance. The design presented here is targeted for applications in which VDD is low during ESD.

The voltage at the I/O pad increases during the rising edge of an ESD pulse, causing the PMOS transistors M1 and M3 to begin pulling up the nodes connected to their drains. The drain of M1 is connected to the gate of M3 and vice versa, forming a feedback loop with contention during the initial stage of the turn-on process. M1 is larger than M3 so that it pulls up its drain node more quickly. Once M1 raises the

voltage at the gate of M2 above the NMOS threshold voltage V_T, M2 turns on and forms a positive feedback loop with M1. The gates of M2, M3, and M4 are pulled up to the pad voltage by M1 and the voltage at the gate of M1 is pulled down to ground by M2.

Once the voltage at the gate of NMOS transistor M4 increases above V_T, M4 turns on and forms a voltage divider with the resistor $R0$. When the output of the voltage divider reaches the switching threshold of the first inverter, the first inverter will switch to low, the second inverter will switch to high and pull up the P-well of the SCR, and the third inverter will switch to low and pull down the N-well. Current is injected into the N-well and P-well, turning on both BJT's simultaneously and triggering the SCR. The trigger voltage is highly dependent on the voltage divider formed by M4 and $R0$; V_{T1} can be increased by increasing the channel resistance of M4 (R_{M4}) and/or by decreasing the value of $R0$. In practice, reducing its channel width W_{M4} increases the channel resistance of M4.

The voltage divider's effect on the trigger voltage is two-fold. First, as the ratio $R_{M4}/R0$ increases, the output of the voltage divider decreases and the I/O pad voltage must be raised higher so that the divider output can reach the switching threshold of inverter 1. However, as the pad voltage increases, the switching threshold of each inverter also rises. The dependence of the switching threshold of inverter 1 on the I/O pad voltage increases the sensitivity of the trigger voltage to changes in the voltage divider.

4.2.6.4 Pulsed Characterization of Dual-Base Control (DBC) Clamp

In [88], circuit simulations were used to illustrate the effect of the voltage divider on the trigger voltage of the protection device for 90 nm logic CMOS technology, this section presents experimental results.

Dual-base triggered SCR's with an SCR width of 50 μm were fabricated in a 0.13 μm BiCMOS technology (gate oxide thickness = 20 A) with a core VDD of 1.2 V and an I/O voltage of 2.5 V. Devices were characterized with a Kelvin transmission line pulse (TLP) system using 100 ns pulses [93].

During an ESD event, the chip is unpowered and thus VDD was left floating during this set of TLP tests. Figure 4.12 shows the pulsed I–V characteristics of three different dual-base triggered SCR's. The pulse rise time was 10 ns.

The three SCR's are identical, but their trigger circuits contain different sizes of W_{M4} and $R0$, thereby altering the voltage division ratio between M4 and $R0$ and, consequently, the trigger voltage of the SCR. V_{T1} values of 3.1, 3.7, and 4.2 V were obtained. These circuits were designed to protect a 2.5 V I/O pad. They were designed to have a V_{T1} greater than the I/O voltage of 2.5 V, because they were designed for an application where a power-on sequence cannot be assured. For other applications, the trigger voltage could be varied over a wider range.

The core VDD line is used as a control signal to reduce leakage and increase the trigger voltage when the chip is powered on. The trigger voltage, V_{T1}, was measured under power-on and unpowered conditions using pulses with 2 and 10 ns rise times. Three different designs were characterized. In the three designs, the transistor sizing

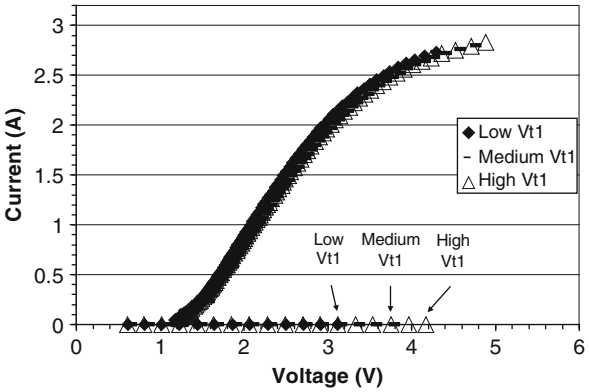


Fig. 4.12 Pulsed *I–V* of three dual-base triggered SCR's designed to have different trigger voltages

Table 4.1 Trigger voltages of three designs with varying inverter sizing ratios

	Design 1	Design 2	Design 3
W_{M6}/W_{M5}	13.6	27.3	41
W_{M8}/W_{M7}	11.4	22.7	34.2
V_{T1} (VDD=0, $t_R = 10$ ns)	3.13 V	3.74 V	3.34 V
V_{T1} (VDD=0, $t_R = 2$ ns)	3.13 V	3.76 V	3.56 V
V_{T1} (VDD=2.5 V, $t_R = 10$ ns)	7.18 V	7.18 V	9.91 V
V_{T1} (powered, $t_R = 2$ ns)	7.00 V	5.71 V	4.00 V

ratios for inverters 2 and 3 were varied. In all designs, $W_{M6} > W_{M5}$ and $W_{M8} > W_{M7}$, which promotes quick triggering action.

Summary of the experimental validation for different design parameters of the control circuit presented in Table 4.1 for the different conditions on VDD node. The data presented for two different rise time parameters of the TLP pulse t_R. V_{T1} is virtually independent of the pulse rise time when VDD is disconnected. Furthermore, V_{T1} is always higher when VDD is powered on than when it is unpowered.

However, the value of V_{T1} under power-on conditions is affected by the pulse rise time. This is a dV/dt triggering effect that can be mitigated with proper design. The magnitude of the dV/dt triggering effect is dependent on the transistor sizing ratios in the second and third inverters, W_{M6}/W_{M5} and W_{M8}/W_{M7}.

4.2.6.5 DC Leakage of DBC Clamp

One of the most critical advantages offered by the solution is the low leakage. To minimize power consumption, small DC leakage is desirable for ESD protection circuits.

Figure 4.13 shows DC *I–V* curves for the second design of the dual-base triggered SCR, with currents normalized per unit SCR width (A/μm) measured at three

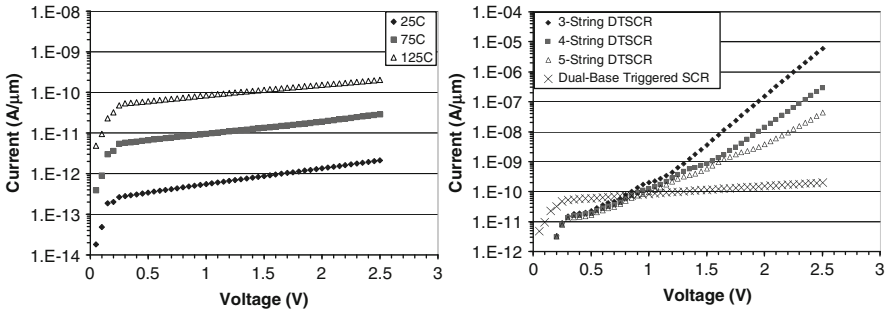


Fig. 4.13 Normalized DC leakage currents for second dual-base triggered SCR design (**a**) and normalized DC leakage currents for three DTSCR's and a dual-base triggered SCR at a temperature of 125°C (**b**) [88]

temperature points. When VDD is supplied, as would be the case under normal operating conditions, the dual-base SCR exhibits flat, low-leakage characteristics throughout the tested range (0–2.5 V). At 125°C, the leakage current is equal to 0.2 nA/μm at a pad voltage of 2.5 V.

Figure 4.13b compares normalized DC *I–V* curves for three DTSCR's with varying diode string lengths fabricated in the same technology with that of a dual-base triggered SCR at a temperature of 125°C. At elevated temperatures, the DC leakage of the dual-base triggered SCR is lower by orders of magnitude than that of the DTSCR's. At a pad voltage of 2.5 V, the normalized leakage of the dual-base triggered SCR is 200 times lower than that of even the 5-string DTSCR with a much higher triggering voltage V_{T1}.

4.3 Voltage and Current Reference in ESD Clamp

Dependence of the ESD clamp performance on the ESD pulse timing characteristics is not always a desirable feature. In case of high capacitive or current load at the pin, the rise time characteristics of the ESD pulse can be significantly altered.

To avoid load-dependent clamp operation, the avalanche breakdown voltage reference can be implemented. The implementation is based on the fast avalanche multiplication process due to a practically local dependence of the multiplication coefficient upon the electric field in Si material. If the process permits change in the blocking junction using the non-self-aligned approach or by additional available implants, then the internal blocking junction can be altered in the snapback device. This measure may provide a desirable effect if the dynamic coupling of the junctions and control electrodes is suppressed on the device level. However, this is not always possible, especially in the case of high-voltage devices, due to a conflict between the dimensions of the required drift regions, minimum process design rules and mask alignment tolerance.

Therefore, in the opposite case, additional voltage reference components can be added to the clamp to provide the desired turn-on level. In this case, understanding snapback ESD devices as three or more terminal components is important. Practical value of the corresponding measurements comes from the dependence of the triggering voltage of the ESD device for different conditions on the control electrode. This information is further used to design appropriate triggering circuit components.

Depending on the type of ESD device, the voltage reference can be implemented relative to the highest potential of the clamp (protected pad) or the lowest potential of the clamp (ground).

The most typical voltage reference component is an avalanche diode. In the case of field electrodes, for example, the polygate in a CMOS structure, the avalanche diode will provide a voltage drop at the breakdown voltage level. In the case of junction-based control electrodes, the avalanche current is used to inject carriers into the electrode.

4.3.1 Low-Voltage Clamps in BiCMOS process technology

Clamp design with a high-side avalanche diode requires an isolated diode structure, which can be used with both the N-type MOS device and NPN BJT ESD devices (Fig. 4.14a, b, c).

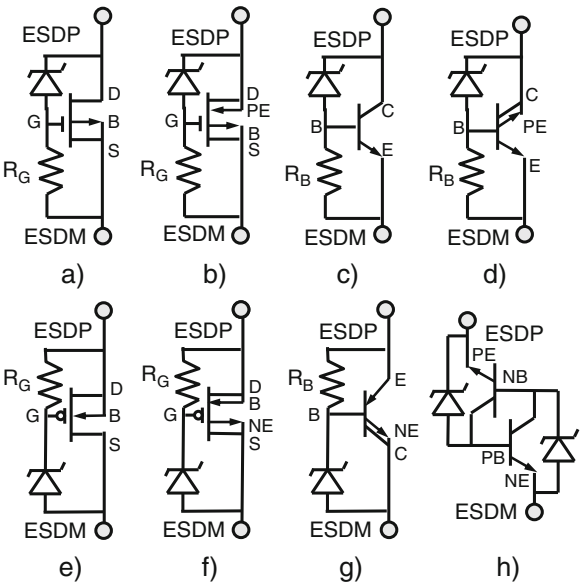


Fig. 4.14 Typical schematics for high-side (**a–d**) and low-side (**e–g**) breakdown voltage reference clamp implementation using avalanche diodes for the case of NMOS and NLDMOS (**a**); NLDMOS-SCR (**b**), NPN (**c**), BSCR (**d**), PMOS or PLDMOS (**e**), and PLDMOS-SCR (**f**), PBSCR (**g**) and the SCR with low and high-side double reference diodes (**h**)

In the case of the P-type device, the low-side avalanche diode can have a p-region that is not isolated from the grounded substrate (Fig. 4.14e, f).

In the case of the SCR device, a similar approach can be applied through decoupling of the SCR blocking junction base region contacts from the injection regions (Fig. 4.14d, g, h).

The principle of operation of the clamp is based upon utilization of the inverse dependence of the triggering voltage upon either the voltage of the field control electrode or the current through the base electrode. In practice, this condition is satisfied rather often (Chapter 3), but not always. Therefore, in order to properly engineer the clamp, a three-terminal characterization of the device (similar to pulsed SOA) is rather important. The purpose of such characterization is to collect information about the instability boundary of the device in order to calculate the clamp parameters.

In the case of CMOS electrodes, this calculation is simple. The current path can be assumed through the clamp resistor during the most part of ESD. Thus, the desired voltage drop on the resistor can be easily estimated. In the case of base junction structures, the total current can be roughly calculated, assuming a base potential of ~0.7 V.

Experimental results for implementation of the avalanche diode reference (Fig. 4.15a) in the NMOS and PMOS clamps discussed above are presented in Fig. 4.15c, d. This design approach provides the lowering of the triggering voltage to a level that corresponds to the sum of the reference voltage and the threshold voltage of CMOS devices. The results provide a practical benefit in adjusting the ESD protection window for these 5 V CMOS devices much closer to the absolute maximum limit of ~6 V.

In principle, any other active device or a stack of components can be used to produce the voltage or current reference for the control electrode of the ESD devices. In the previous section, GGSCR and DTSCR have already been mentioned. The BVCEO-referenced BJT clamp includes a small reference NPN with an open base that is the source of the base current for the main snapback ESD NPN device in their common emitter circuit.

Finally, the avalanche diode can form a useful clamp by itself, as in the example of the avalanche diode I/O and power clamp is presented in Fig. 4.16. However, the capability of such a solution is significantly limited by the ~0.1 mA/μm current density that can be obtained in avalanche breakdown conditions.

This limitation is especially severe in high-voltage devices, where the lengths of the blocking junction regions result in additional current reduction. An example of the avalanche diode clamp with a 25 V voltage tolerance is presented in Fig. 4.17. While the clamp itself can withstand relatively high current, the useful voltage reference and clamping can only be achieved in a rather narrow voltage range with a current of 1–3 mA.

These results, however, are not negative. The major practical application of the avalanche diode as a clamp belongs in two-stage protection circuits. This approach will be discussed in Chapter 5.

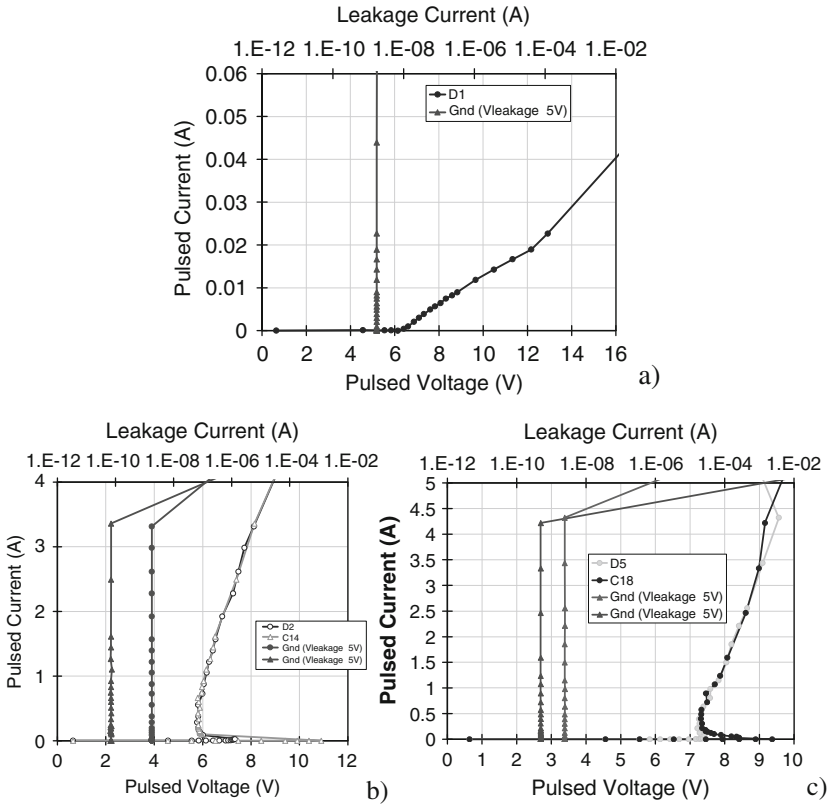


Fig. 4.15 *I–V* characteristics of the avalanche diode (**a**) and comparison of the triggering characteristics for the transient triggered and voltage referenced NMOS (**b**) and PMOS (**c**) snapback clamps

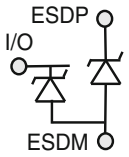


Fig. 4.16 Example of utilization of the avalanche diode as an input and power clamp

4.3.2 NPN Clamps with Voltage Reference

This section compares NPN BJT clamps with different triggering techniques. The overall purpose of this particular design is to implement an 8 V ESD protection in a low-cost isolated junction BJT process [94, 95].

Comparative analysis between transient triggered and different voltage referenced ESD power clamps is presented in [96]. Several architectures of ESD clamps

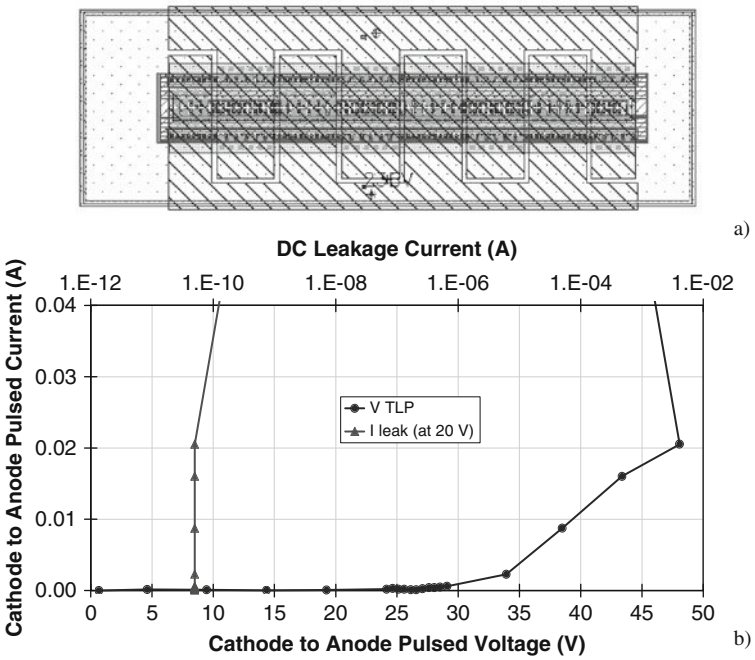


Fig. 4.17 Example of the layout (**a**) and pulsed *I–V* characteristics (**b**) for 25 V reference voltage lateral avalanche diode

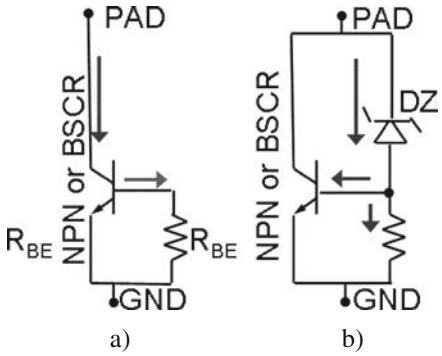


Fig. 4.18 Negative and positive base current directions in the BVCER (**a**) and enhanced avalanche (**b**) BJT ESD clamps

with both external and internal breakdown voltage reference techniques are compared. In particular, the grounded base or BVCER BJT clamp (Fig. 4.18a) is compared to an enhanced Zener clamp (Fig. 4.18b). In spite of a similar common emitter circuit, the operation of these clamps is different, relying on different direction of the base current (Fig. 4.18).

The turn-on voltage for conductivity modulation mode of the first clamp is generally based upon either the breakdown of the internal collector-base blocking junction

or the transient dV/dt effect related to the displacement current in the collector-base junction.

The second clamp is specifically designed to be less sensitive to the d*V*/d*t* effect due to the fast avalanche breakdown in a relatively small avalanche reference diode in the base–collector circuit. Thus, turn-on of the BJT into conductivity modulation mode occurs at negative and positive base currents for the first and the second cases, respectively.

The avalanche breakdown voltage reference can be achieved by using avalanche diodes formed by the PBASE-NEMITTER and PBASE-NSINKER lateral junctions (Fig. 4.19).

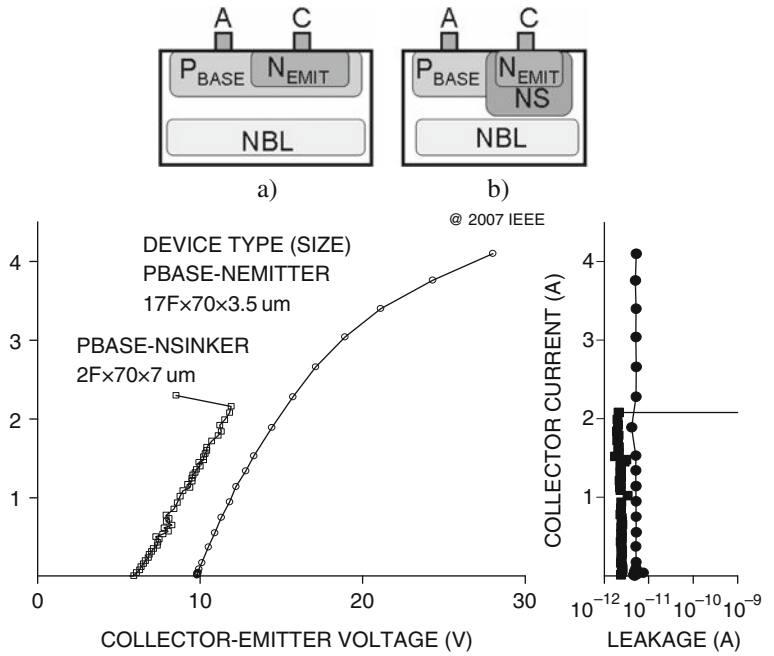


Fig. 4.19 Simplified cross-sections and experimental TLP characteristics for the PBASE-NEMIT (**a**) and PBASE-NSINK (**b**) lateral avalanche diodes [96]

Similar experiments have been performed with ESD clamps using bipolar SCR (BSCR) [7, 97] ESD devices.

Another version of the ESD clamp device has been designed with the internal avalanche diode. This internal junction is formed by stretching the PBASE diffusion to create an overlap with the collector region, thus forming a surface junction between the base and the N-sinker regions (Fig. 4.20). Similarly, the internal avalanche diode structure is implemented in the BSCR, where an additional floating N-EMIT region is added between the PBASE and the P-EMITTER.

Comparison of TLP characteristics for different NPN BJT clamps is presented in Fig. 4.20. In the case of the clamp with external avalanche reference, the

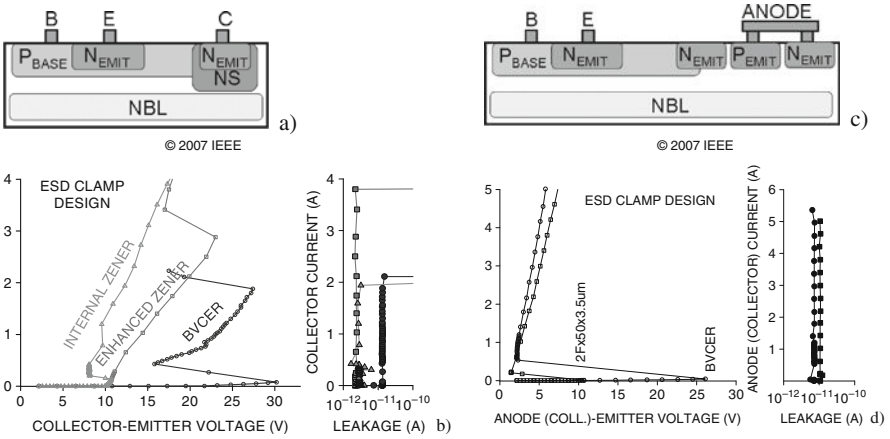


Fig. 4.20 Avalanche breakdown voltage reference effect in NPN BJT and BSCR clamps. Cross-sections for NPN (**a**) and BSCR (**c**) with internal voltage reference; measured TLP *I–V* characteristics for the BVCER, enhanced Zener clamp and internal avalanche NPN clamp (**b**) and for internal Zener NPN and BSCR clamps (**d**) [96]

breakdown voltage of the device is significantly reduced, the critical current required for snapback is rather large, and device operation is not optimal in terms of providing the proper waveform in the desired voltage range (8–10 V for the particular application). In contrast, the device with the internal blocking junction provides excellent triggering characteristics that provide both the low triggering current and the low holding voltage.

Similar snapback characteristics, but with better high-current tolerance, were observed in the BVCER and internal Zener BSCR clamps (Fig. 4.20). Also, in the case of the base and N-sinker overlap, the corresponding blocking junction was formed with a lower breakdown voltage of ~7 V.

In spite of the similarities in design and operation, the internal and the external voltage reference clamps have quite different triggering mechanisms. These mechanisms are defined by the direction of the base current and the interplay between the avalanche and the injection effects. These effects were already discussed in Chapters 2 and 3.

It should also be taken into account that from the structural point of view, the NPN BJT is composed of vertical and lateral NPN devices. At lower current values, the vertical BJT is involved in a conductivity modulation mechanism, thus providing the initial part of the snapback characteristic. The lateral surface NPN is engaged after subsequent current increase.

4.4 High-Voltage ESD Devices

As was mentioned before in this book, a conventional submicron process subdivision between low and high-voltage pins is assumed at the level of ~12 V. In this

section, clamp design specifics for high-voltage applications are discussed for the BiCMOS and extended MOS processes in the 20–100 V range.

4.4.1 20 V NPN with Blocking Junction Internal Reference

An example of the 20 V NPN clamp with an internal voltage reference is presented in Fig. 4.21. In this cell, a low-voltage Nwell implant has been used to reduce the breakdown and triggering voltage from the original standard device level of ~35 V down to the 28 V. This margin is appropriate for pins with an operation voltage of 20 V, taking into account the temperature dependence and the statistical deviation of standard device SOA parameters.

The clamp has practically no d*V*/d*t* sensitivity in the 2–10 ns rise time domain.

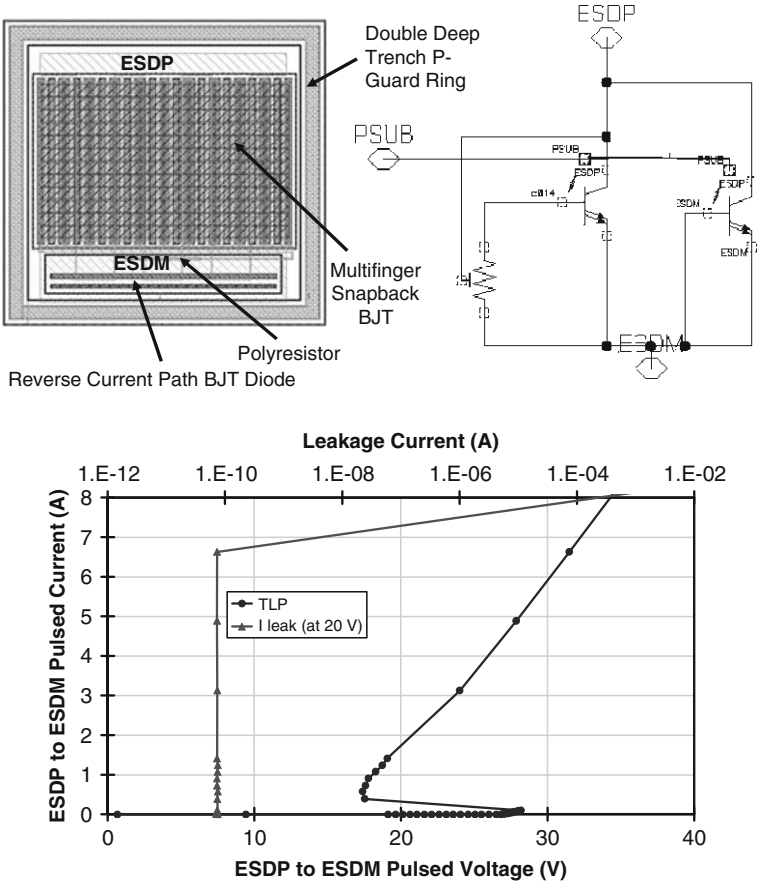


Fig. 4.21 High-voltage avalanche breakdown voltage reference NPN BJT clamp layout and schematic view and measured TLP characteristics

To enable automatic voltage tolerance for the high-voltage reverse-path diode component, an additional shorted base NPN structure is used. In this case, the collector–emitter diode provides the reverse pulse protection in negative ESD pulse. An alternative way to build a more compact cell is to short some base–emitter junctions in a few selected fingers of the snapback BJT. In this case, it is expected that the fingers with shorted base, dedicated to reverse current path protection, will also contribute into the total collector current in the snapback mode.

One of the important considerations of clamp design is to take into account the high-voltage isolation of and interaction with the guard ring structure.

4.4.2 NPN Clamp with External Lateral Avalanche Diode Reference

An alternative example of the external avalanche diode 20 V NPN clamp implemented in the same 20 V BCD process with external voltage reference is presented in Fig. 4.22.

In this case, the reference avalanche diode component is the same as the one presented in Fig. 4.17.

Similarly to the clamp in Fig. 4.21, this clamp also has practically no dV/dt sensitivity in the 2–10 ns rise time domain.

4.4.3 SCR-Based High-Voltage Clamp

An alternative clamp can be created using SCR-type ESD devices (Fig. 4.23). For example, a high-voltage BSCR can be obtained from the NPN device. Similarly, an NLDMOS-SCR can be obtained by a corresponding transformation of the standard 20 V NLDMOS supported in the process. Both structures require embedding of the p+-emitter region in order to form a parasitic p–n–p structure, in addition to the n–p–n structure already present in the original device (Chapter 3).

The parasitic PNP can be formed between the p^+ guard ring, p^+-emitter, and NDRIFT or collector epi isolation (acting as a base).

A similar clamp can be composed for the NDeMOS-SCR implemented in high-voltage extended-drain devices.

4.4.4 Lateral LPNP Clamp

High holding voltage protection using lateral PNP structures has already been discussed in this chapter. The implementation of the clamp mainly targets the high-density layout to achieve maximum possible total width of the device (Fig. 4.24). For this purpose, base contacts are assembled using the minimum diffusion region at the finger side.

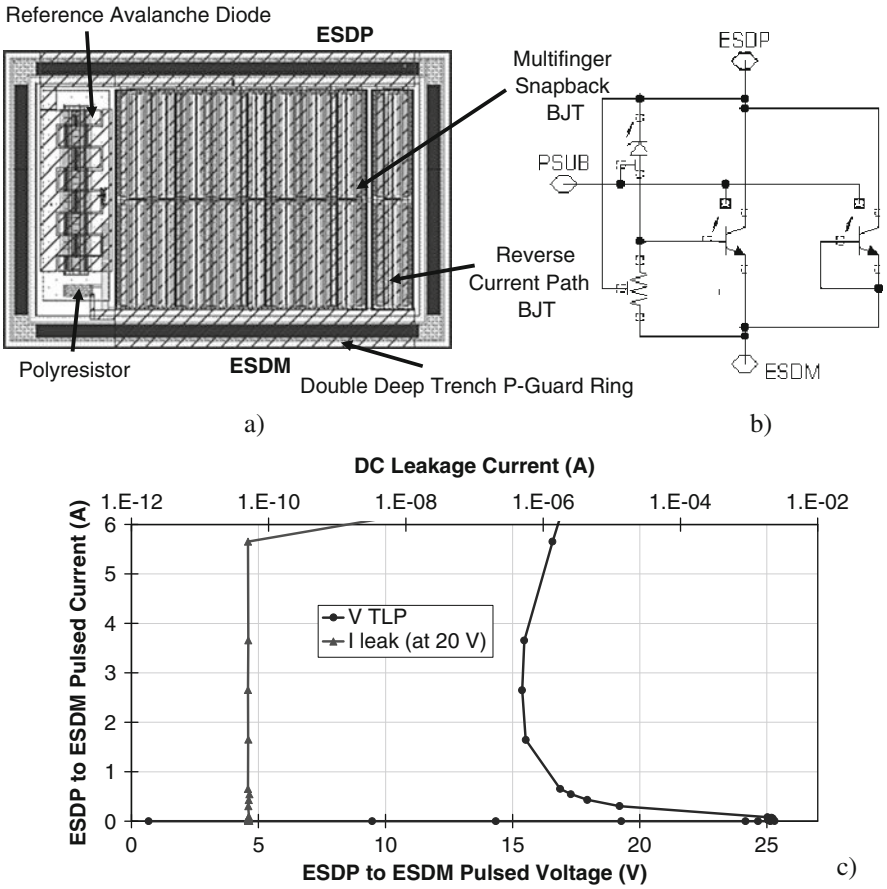


Fig. 4.22 High-voltage NPN BJT clamp with collector-emitter avalanche diode reference: layout view with clamp components (**a**), schematic view (**b**), and measured TLP characteristics (**c**)

One of the specifics for clamp implementation is the n-base connection. The n-base connection is necessary to eliminate excessive base-collector coupling that can be undesirable for pins with high transient voltages.

4.4.5 Mixed Device-Circuit Dual Mode Solutions

The basic principle of the mixed-mode device-circuit solution has already been demonstrated for the dual-base-controlled SCR in Section 4.2.5.

In this section, a similar approach is illustrated for the high-voltage case, using the example of a 50 V clamp [85, 89, 90].

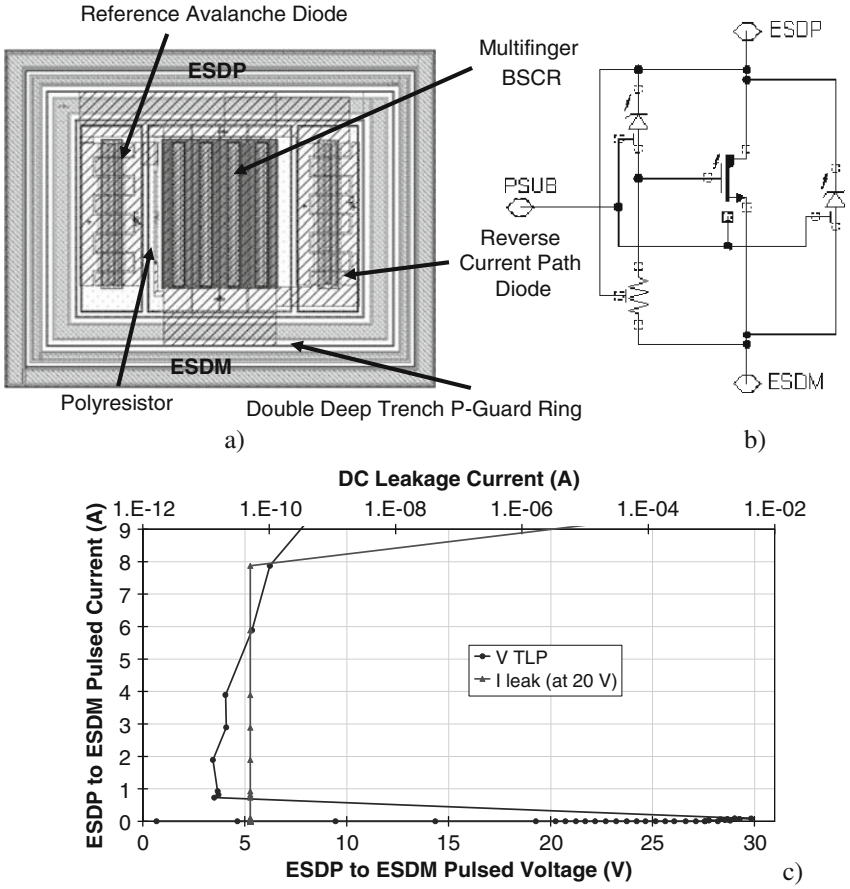


Fig. 4.23 Example of the 20 V NLDMOS-SCR layout, schematic view, and TLP characteristics

The trend in improving switching capabilities of high-voltage devices usually results in a rather small or even negative ESD protection window, as a side effect of the technology improvement.

Thus, application of both voltage referenced and transient triggered clamps becomes undesirable and rather challenging, from the point of view of achieving the required parameters. In this section, an alternative way to implement ESD protection is discussed with the purpose of demonstrating the physical concept of using ESD devices as active structures with snapback characteristics that are actively controlled by the circuit driver.

4.4.5.1 Example of Circuit Design

A simplified schematic of the active control circuit is presented in Fig. 4.25. The circuit contains a driver that controls the gate of the snapback device MESD and

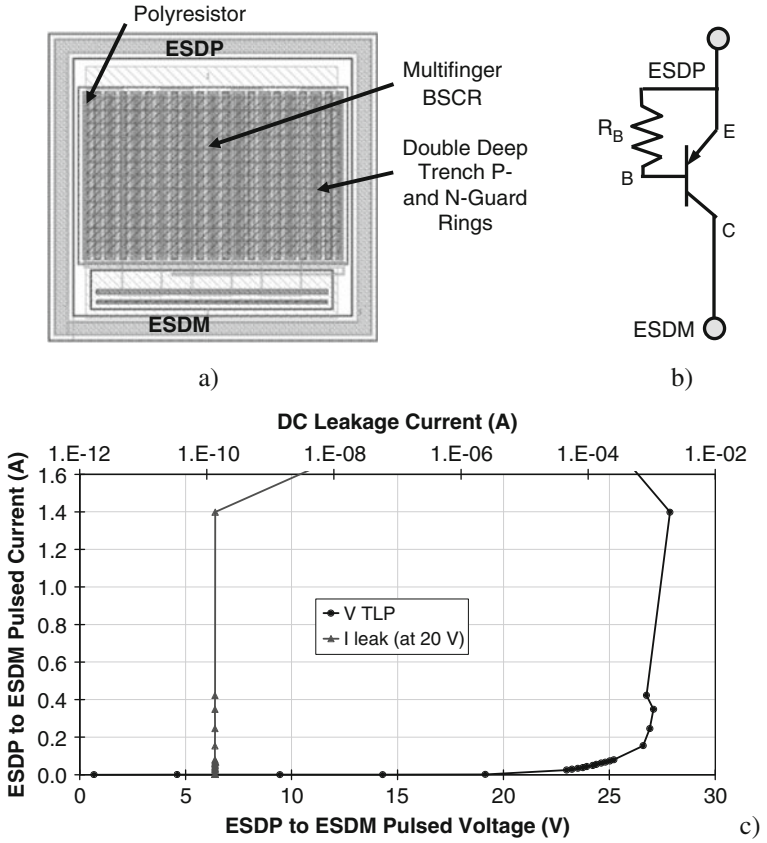


Fig. 4.24 Example of the 20 V lateral PNP layout, schematic view, and TLP characteristics for the NLDMOS-SCR

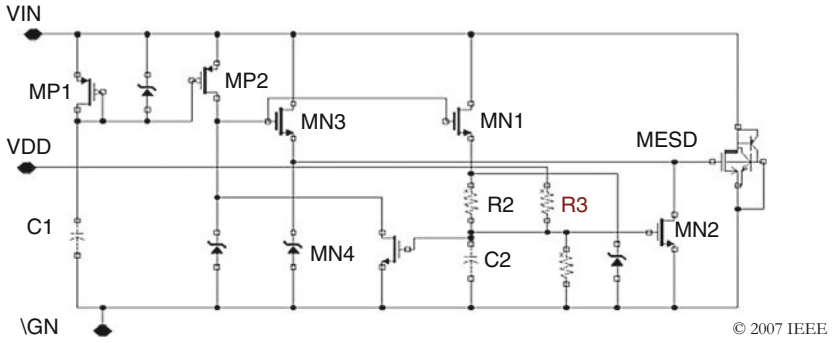


Fig. 4.25 Active driver circuit with snapback SCR device (MESD) controlled from the gate terminal [85, 89]

operates depending on the signal from the VIN and VDD pins. This circuit has been studied using circuit simulation.

In the case of a positive ESD event on the VIN pin, MP2 will be on until the capacitor C1 is fully charged (∼20 ns). During this time, MP2 and MN3 directly apply a positive bias on the MESD control electrode (gate). This creates the conditions for fast triggering of the MESD into snapback at rather low VIN pad voltages. Significantly, the circuit does not only depend on the capacitive coupling to pull the gate high, as in some alternative designs. MP2 also turns on MN1, which sources current for a turn-off RC-circuit composed of resistor *R*2 and capacitor C2. After a corresponding RC-delay of ∼200 ns, the transistors MN2 and MN4 turn on. MN2 shorts the control electrode of the ESD device to ground, while MN4 turns off both MN3 and MN1. MESD remains in high-current mode until the ESD discharge is completed. The low-voltage avalanche diodes are used to limit the gate voltage of the MOS devices. During an ESD event, the VDD potential is near ground due to the high capacitance and leakage of internal circuitry of the voltage regulator.

In the case of normal operation, during the relatively slow (determined by the total equivalent capacitance of the chip and the external devices) powering sequence of VIN above ∼5 V, the VDD signal is generated in the internal circuitry; MN2 is then turned on through *R*3. MN2 holds the control electrode low. This maximizes the trigger voltage of MESD. The VDD signal will also turn on MN4. MN4 will keep MN3 and MN1 off, regardless of VIN, eliminating the chance of contention between MN3 and MN2. Holding these transistors off also provides for low leakage in the clamp circuit.

This ESD solution has been experimentally validated in [89] for a 0.35 μm CMOS process with 20 V extended drain MOS devices (DeMOS). For pulses with a 10 ns rise-time, V_{ABSMAX} is 27 V. Given $V_{\mathrm{MAX}} = 22$ V, the protection window is 22–27 V (Fig. 4.26a). With a 200 ps rise time, the upper limit is lowered to ∼23 V. The DeMOS-SCR lateral ESD protection device used is similar to the NLDMOS-SCR device (Fig. 4.26a).

The TLP *I*–*V* characteristics for two different rise-times and under two different operating conditions, at VDD=0 V (the results were the same as for a floating VDD node) and at VDD=5 V, are presented in Fig. 4.26b. Under the condition that corresponds to ESD stressing of the non-powered circuit (VDD=0 V), the triggering voltage is very low (∼5 V). This guarantees reliable ESD protection not only for high-voltage devices but also for the low-voltage devices in the internal circuit that are connected to the protected pin. However, the leakage measured at VIN=20 V remains the same as in a grounded gate ESD device (Fig. 4.26b, left plot).

In the case that corresponds to normal operation (VDD=5 V), the trigger voltage of the clamp is high (∼29 V) and equivalent to that of an NLDMOS-SCR device with a small gate-source resistor. Evidently, the clamp may be designed to have a triggering voltage that, under normal operation, exceeds V_{ABSMAX}. This is desirable, since a high trigger voltage correlates to unreliable snapback operation and can result in damage of the devices. ESD operation of the clamp is fully controlled

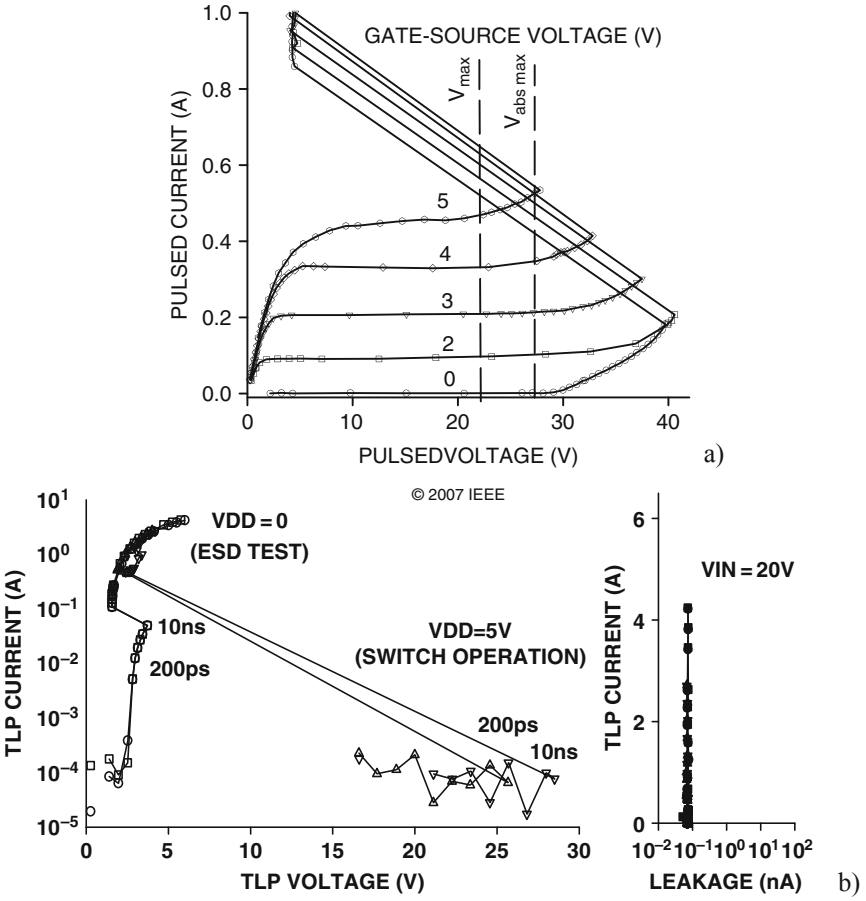


Fig. 4.26 Pulsed output characteristics and ESD protection window for the 25 V absolute maximum voltage DeMOS device (**a**), and measured TLP results for the mixed device-circuit solution clamp showing both modes of operation (**b**) [89]

by the channel current, rather than the avalanche breakdown current, thus always providing reversible snapback operation.

It has been shown that both overshoot and turn-off voltages can be controlled by this solution (Fig. 4.27). This example demonstrates a cross-disciplinary aspect of future advanced physical ESD design. The ESD device, as a component of the dual mode clamp, should be designed with a proper dc voltage tolerance in order to provide low sensitivity to the fast rise time and, at the same time, afford sufficient control of the triggering characteristics (reversible instability boundary) by the control electrode, with reversible high-current operation and a small footprint.

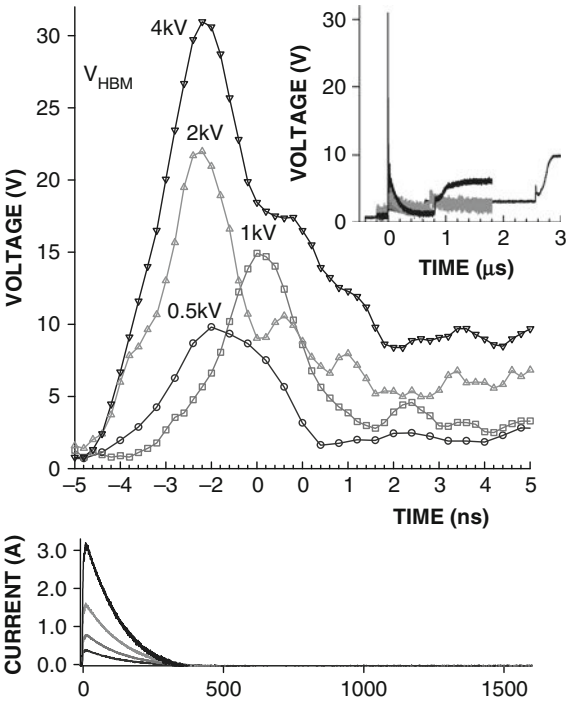


Fig. 4.27 Comparison of the measured voltage waveforms for the active gate control circuit [89]

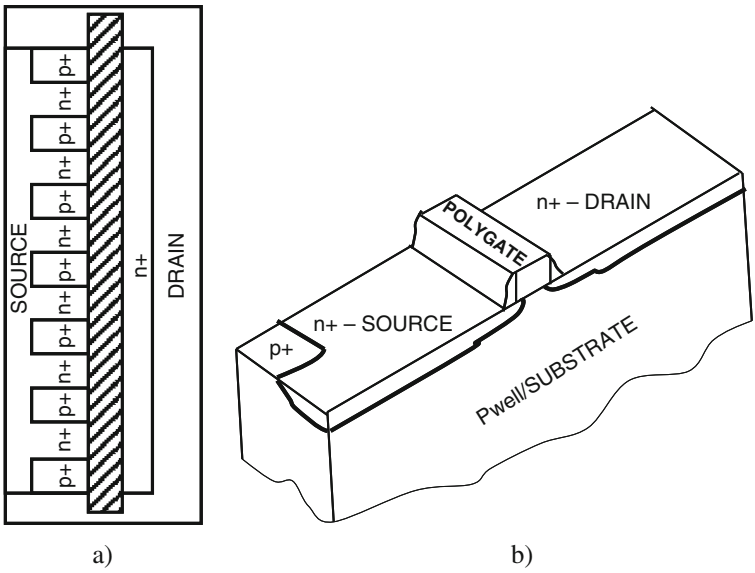
4.5 The Concept of Self-Protection

Self-protection capabilities have been already mentioned indirectly in the discussion of SOA for standard devices. In the most simplified case, if the devices connected to the output pin are of adequate size, the pin will not require an additional clamp. In a more general case, the self-protection capability can be achieved at different levels: the device level, circuit level, or array level.

4.5.1 Device-Level Self-Protection

A number of changes can be made in the architecture of the standard device in order to implement self-protection capabilities.

One of the most common examples of self-protecting solutions is the addition of the drain ballasting region. With simultaneous increase in the well tap spacing, I/O NMOS devices can reversibly withstand the snapback operation similarly to ESD clamps. In normal operation mode, the side effects of this design are an increased space on the chip, parasitic drain capacitance, and a higher on-state resistance.



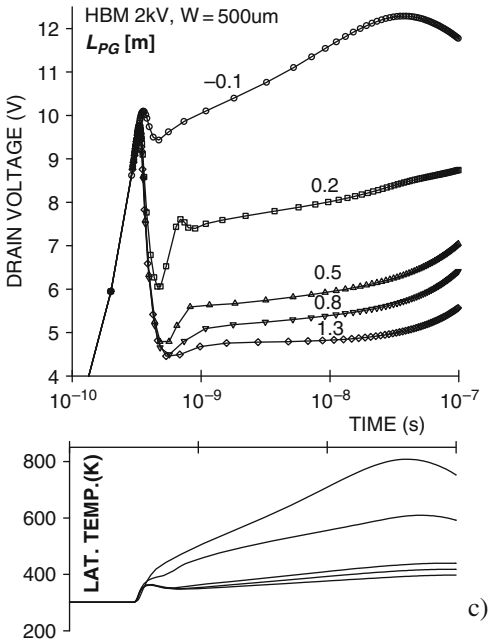


Fig. 4.28 Power array finger with interdigitated source/P-well diffusions (**a**), 3D equivalent structure (**b**), and the results of 3D numerical simulation analysis (**c**) of the waveforms of mixed-mode HBM simulation upon the distance between the PPLUS region and the edge of polygate (L_{PG})

Another way to implement an additional self-protection capability or at least push the current instability level to a higher voltage and current range is to use an interdigitated well tap connection solution (Fig. 4.28a). In this case, the parasitic NPN bipolar has the minimum possible parasitic base (bulk)–emitter (source). This measure provides the lowest current gain in the internal parasitic bipolar structure for NMOS devices. Since the avalanche–injection conductivity modulations occur at a much higher critical avalanche current, both the instability boundary and the ESD protection window are changed.

This solution is especially practical in fully silicided processes that provide a reliable connection to the interdigitated regions with no special contact placement.

From the physical point of view, operation of this device is based upon reduction of the hole density in the discharge region due to a reduced gain of the parasitic NPN BJT, and effective extraction of the carriers by the p+-contact regions. Respectively, the average electric field amplitude under the gate in the triggered-on state is increased, due to a lack of space charge compensation between the electrons injected from the source junction and the holes generated in the avalanche multiplication drain region. In particular, this results in an increase of the holding voltage. TCAD-level validation for this solution is presented in Fig. 4.28b, c. The major parameter for comparative analysis in this case is the distance L_{PG} between the introduced PPLUS region and the source side of the polygate region.

The mixed-mode signal waveforms (Fig. 4.28b) HBM ESD operation demonstrates that the holding voltage realized for the structure can be even bigger than the triggering voltage. Analysis of the 3D electron and hole current distributions exhibits the conductivity modulation process and the holding voltage increase in greater detail.

A similar solution has been proposed for high-voltage lateral DMOS devices, where the P-buried layer is used to reduce the corresponding internal base–emitter resistance and common base current gain of the parasitic lateral BJT, thus improving the SOA of the device.

4.5.2 Array-Level Protection

A more sophisticated concept is proposed in [98], where self-protection capability of the arrays is achieved by embedding the additional parasitic SCR structure inside the array. In this case, an embedded ESD device is engaged to take over ESD current conduction, before limitations in the parasitic NPN device of the standard array occur due to avalanche–injection. Contrary to a conventional local clamp approach, this concept provides the self-protection capability within the array itself, since the area of the embedded ESD device also contributes to the on-state current during normal operation.

This solution was experimentally demonstrated for an NLDMOS power array in a 50 V BiCMOS process. Series of distributed p+-emitter regions and diffusion regions have been embedded into some array fingers in order to form an additional parasitic SCR structure with reversible snapback capabilities.

The self-protecting array (SPA) capabilities have been achieved by transformation of small parts of the NLDMOS array into an NDMOS-SCR that provides current conduction during the ESD event. At normal operation, the NDMOS-SCR portion of the device contributes toward the useful array current.

The equivalent circuit of the self-protecting array device is presented in Fig. 4.29d, where the left part represents the fingers of the NLDMOS and the right part represents the NLDMOS-SCR.

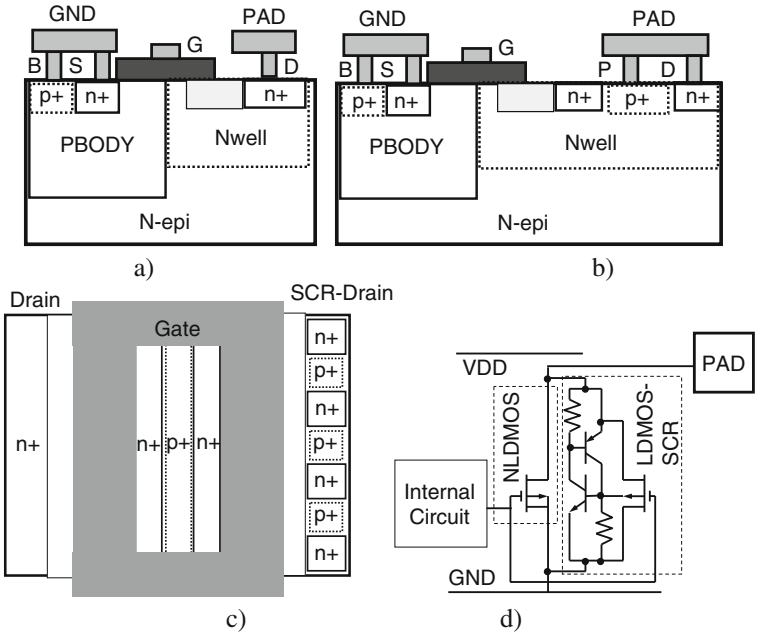


Fig. 4.29 Simplified cross-section for NLDMOS (**a**), local clamp NLDMOS-SCR devices (**b**), and topology (**c**) for the self-protecting device with the left finger formed by a standard supported device and right finger formed by the embedded SCR device; (**d**) simplified circuit diagrams for the self-protecting array [98]

Simplified cross-sections of the NLDMOS and NLDMOS-SCR devices and the constructed topology of the self-protecting device, based upon the principles listed above, are presented in Fig. 4.29a, b. The topology is utilized by additional interdigitated P-emitter regions added into the drain composite regions of one or more device fingers, depending on the array size (Fig. 4.29c).

These experimental SPA devices provide different pulsed *I–V* characteristics when compared to standard control NLDMOS array devices (Fig. 4.30). Unlike standard devices, the SPA's are able to sustain a larger current level suitable for reversible ESD protection. At the same time, major figures of merit associated with the array's normal on-state operation have been changed negligibly with acceptable change in the pulse ESD SOA of the device itself (Fig. 4.31).

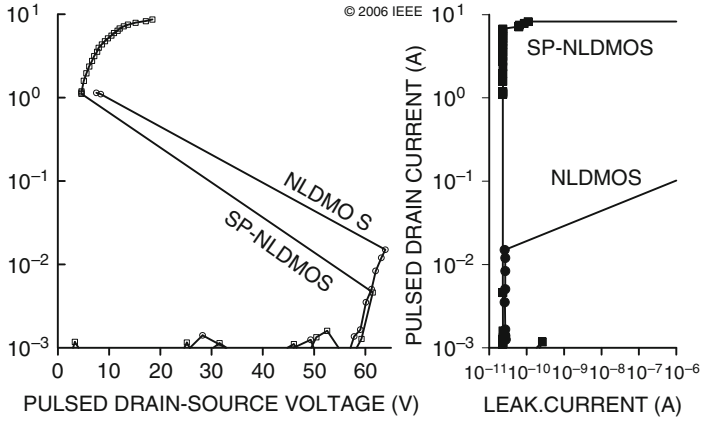


Fig. 4.30 Comparison of TLP characteristics for the standard and self-protecting 50 V NLDMOS array with a 400 μm width in the grounded gate configuration [98]

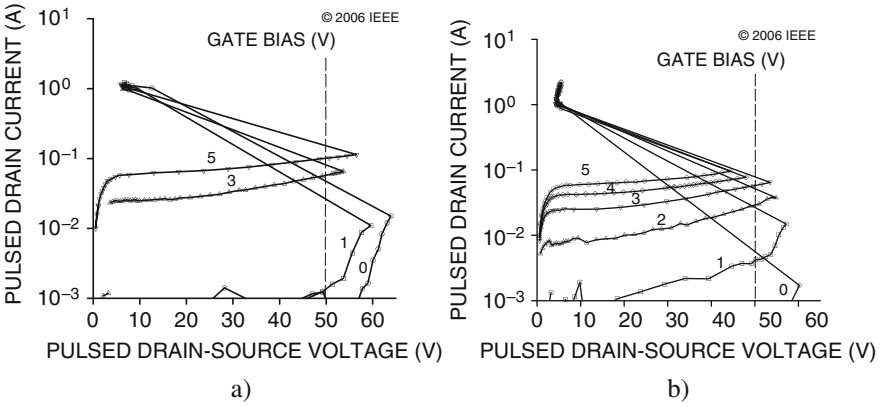


Fig. 4.31 Comparison of the pulsed SOA using output pulsed drain-source TLP characteristics measured at different gate bias for standard NLDMOS (**a**) and SPA arrays (**b**) of similar footprints [98]

This solution is essentially made possible by the implementation of the instability boundary for NLDMOS-SCR below the pulsed SOA limits of the NLDMOS device. This implementation principle generally requires a precise embedded SCR design, where the snapback voltage is a much stronger function of gate bias. At the same time, one of the objectives of the solution is to preserve the SOA of the whole array at the original level.

4.6 ESD Protection of Ultra High Voltage Circuits

Evolution of certain power applications creates the necessity of integrating very- and ultra-high-voltage (UHV) level components on the chip [99–101], up to the

range of 500–1000 V. This voltage level is required for compatibility of the different power sources with the power grid. Meantime the experience in such high-voltage protection is very limited [102, 103].

Recently, an interest has emerged in implementing integrated grid-tolerant circuits. In this application, integrated circuit pins have to be tolerant to 700–1000 V, with some galvanically isolated pins requiring additional system level protection with the voltage tolerance level up to ∼5 kV. One of the typical applications is the solar energy conversion. In this case, the dc voltage level generated by the solar panels needs to be converted to the standard ac current of the grid. These actively developing photovoltaic (PV) applications require circuits with voltage tolerance of up to 800V at current levels of ∼1 A.

Other ultra high-voltage applications include factory automation, electric motor control, and LED lighting.

An example of a photovoltaic system is presented in Fig. 4.32. The left part of the circuit presents a DC/DC converter deriving input voltage from the PV cell in the range of 23–38 V and producing a maximum output voltage of ∼400 V. The DC/DC converter includes a high-frequency DC–AC converter block, a 1:10 transformer providing the AC voltage increase up to 400 V, a diode rectifier, and an LC filter. Then obtained high DC voltage is used by unfolding the DC/AC inverter to create a grid-compatible AC output. The key components are four discrete devices used in the T-bridge with a voltage tolerance over 700 V. The 700 V voltage tolerance is required to achieve the 220 V root mean square (RMS) voltage level. It can be assumed that future systems will require ESD protection of pins with such a high-voltage tolerance.

Perhaps one of the first impressions is that UHV pins may not require ESD protection at all, since the voltage tolerance is already rather high. Nevertheless, according to a more accurate evaluation, ESD protection is still required, especially at the system level.

In spite of their high-voltage tolerance levels, such pins may require ESD protection.

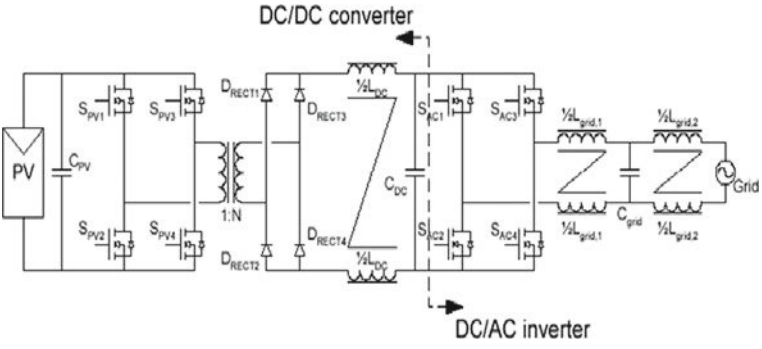


Fig. 4.32 Power circuit of the photovoltaic inverter with UHV components

However, even in such "exotic" cases the basic principles still apply. The most straightforward solution is the implementation of the LDMOS-SCR structure with the p+-emitter isolated by the drain region.

An example of the layout of the 500 V device is presented in Fig. 4.33 with the UHV pad located in the middle of the structure. The typical device topology supports a significantly different scaling for the drift region length (~20–50 μm) in order to control the triggering voltage.

The measured device demonstrates reversible snapback behavior (no failure). The device demonstrates switching into "SCR-mode" after turn-on. Experimental evaluation of the UHV devices presents a serious challenge. It requires significant modification of the high-voltage high-speed voltage probe for waveform evaluation.

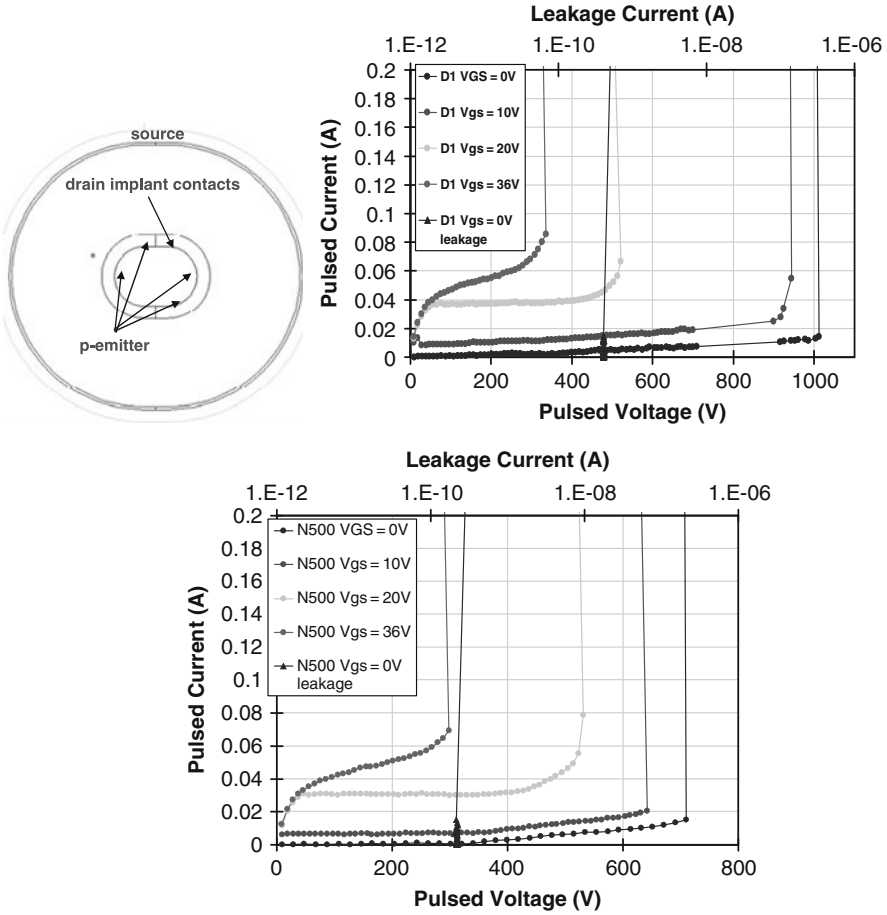


Fig. 4.33 Example of the 500 V device layout with middle pad, TLP characteristics of the standard device, and ESD clamp at different gate bias levels

Another problem is related to a significant voltage and current overshoot observed which increases with precharge level (Fig. 4.34).

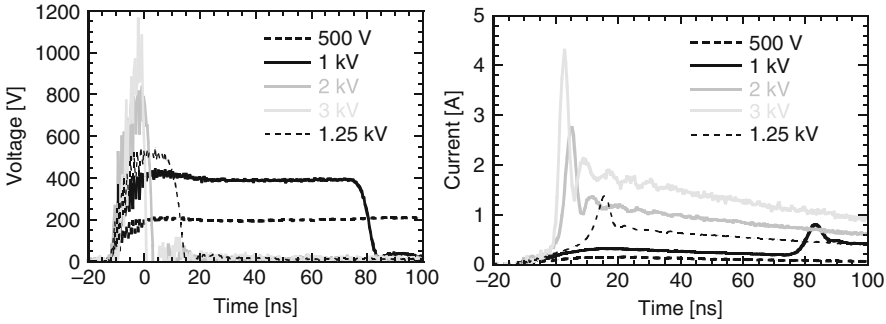


Fig. 4.34 Voltage and current waveforms for 0.5, 1, 2, and 3 kV for 400 V grounded gate NLDMOS-SCR clamp

4.7 Summary

In this chapter, the knowledge of ESD devices and structures (Chapters 2 and 3) has been applied to describe ESD clamps. Implementation of ESD devices, studied through their cross-sections, into distributed layout objects on the semiconductor wafer is not always a straightforward task.

Clamp implementation includes combination of a broad range of aspects: device width scaling, lateral isolation, vertical isolation from the substrate, latch-up guard rings and, of course, all the clamp components for voltage reference, and dynamic coupling.

An emphasis is placed on the methods and possibilities used to achieve the desired voltage waveforms for ESD pulse operation. This has been accomplished in two ways: by the clamp circuit design and by the clamp topology design.

Features similar to the positive and negative feedback effects realized on the device structure level can be engineered at the clamp level as well, in order to control the waveforms produced by the clamp during ESD event.

One of the most effective clamp designs creates voltage and current reference circuits using clamp components in the fast avalanche breakdown mode, as well as more complex circuits for the control of clamp triggering-on and triggering-off.

Layout implementation of the ESD device is a very critical part of the clamp. The topological degree of freedom is advantageous in engineering appropriate characteristics of the clamp itself.

The complexity of the aspects involved in clamp design requires verification and monitoring of the clamp characteristics using extensive test chip experimentation.

DECIMM^{TM} Simulation Examples for Chapter 4

To download a trial version of the numerical simulation software and request an electronic license key please visit http://www.analogesd.com
To download libraries with simulation examples for this chapter please visit http://www.analogesd.com/Chapter4.html
List of examples is subject to change.

Example 4.1 Snapback NMOS Clamp Operation Analysis

Library Name: Examples4_Snapback_Clamps
Project Name: E4.1_5 V_Snapback_NMOS_HBM

This example explores basic operation and optimization of one of the most popular ESD solutions – the snapback NMOS clamp. Comparative analysis of the HBM pulse waveforms can be performed for a range of parameters of the Snapback NMOS device, gate-source resistor *R*2 values and different HBM pulse amplitudes. In particular, the increased gate-coupling effect due to the increase of gate resistor (*R*2) value results in a reduction of the voltage overshoot (Fig. E4.1). Reduction of the snapback device (M3) width results in an increase of the turn-off time (Fig. E4.1). This example also demonstrates the change of the triggering characteristics for different substrate-to-source spacing. It is also important to

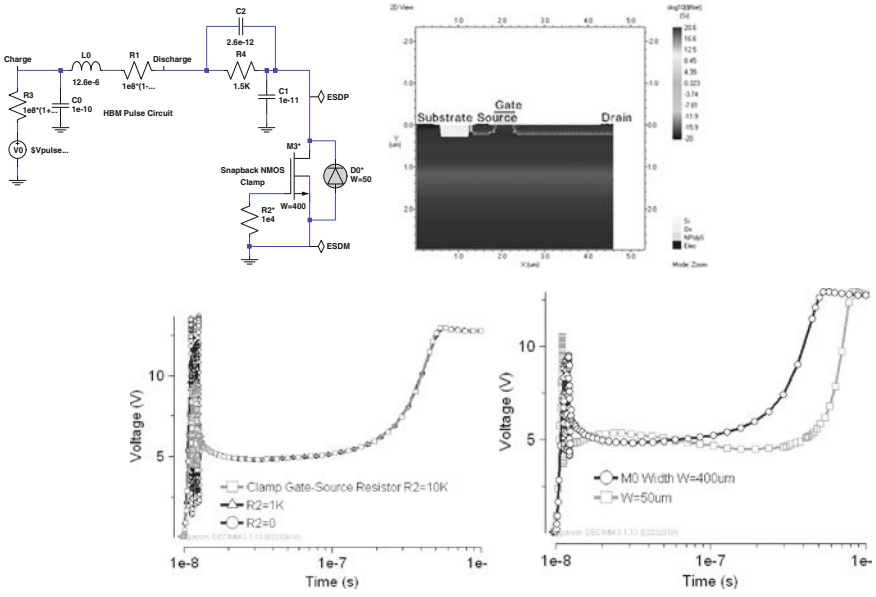


Fig. E4.1 Mixed-mode simulation circuit with a cross-section of the snapback NMOS device and 2 kV HBM pulse response for different gate-source resistor *R*2 values and sizes of the Snapback NMOS device

compare the operation of the Snapback NMOS with that of other snapback devices presented in this section, for example the LVTSCR devices presented in Example 4.2.

Example 4.2 LVTSCR ESD Clamps

Library Name: Examples4_Snapback_Clamps
Project Name: E4.2a_NLVTSCR_HBM
Project Name: E4.2b_PLVTSCR_HBM

These mixed-mode simulation examples study a circuit with parameterized structures for N-type LVTSCR and P-type LVTSCR. The NLVTSCR example is

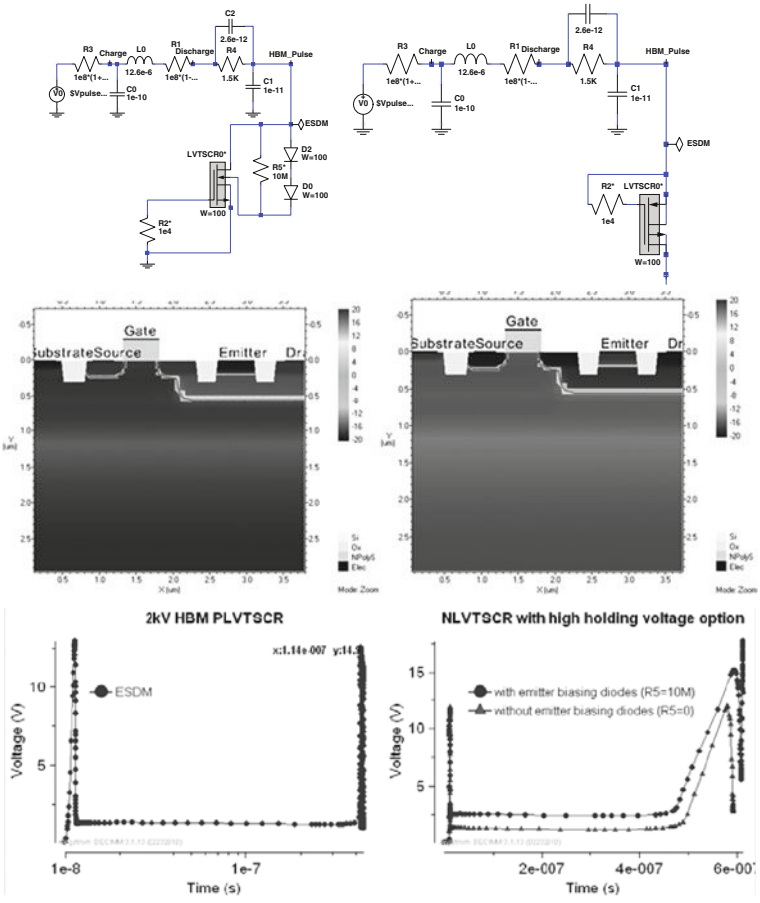


Fig. E4.2 Mixed-mode simulation circuits with cross-sections of NLVTSCR and PLVTSCR and mixed-mode simulation results of the HBM pulse response for different circuit and device parameters

constructed with optional emitter biasing diodes to demonstrate the effect of the holding voltage increase. The effect is used to eliminate the possibility of latch-up by increasing the holding voltage of the SCR above the power supply level. The effect of the diodes is introduced or eliminated by changing the value of the resistor *R*5 in the circuit. A comparison of the HBM pulse waveforms for cases with and without emitter biasing diodes is presented in Fig. E4.2.

Example 4.3 Two-Stage ESD Protection with Snapback NMOS

Library Name: Examples4_Snapback_Clamps
Project Name: E4.3_5 V_Two_stage_HNMOS_HBM

This example demonstrates the two-stage clamp ESD protection principle. The example is composed of a grounded gate snapback NMOS clamp, second stage

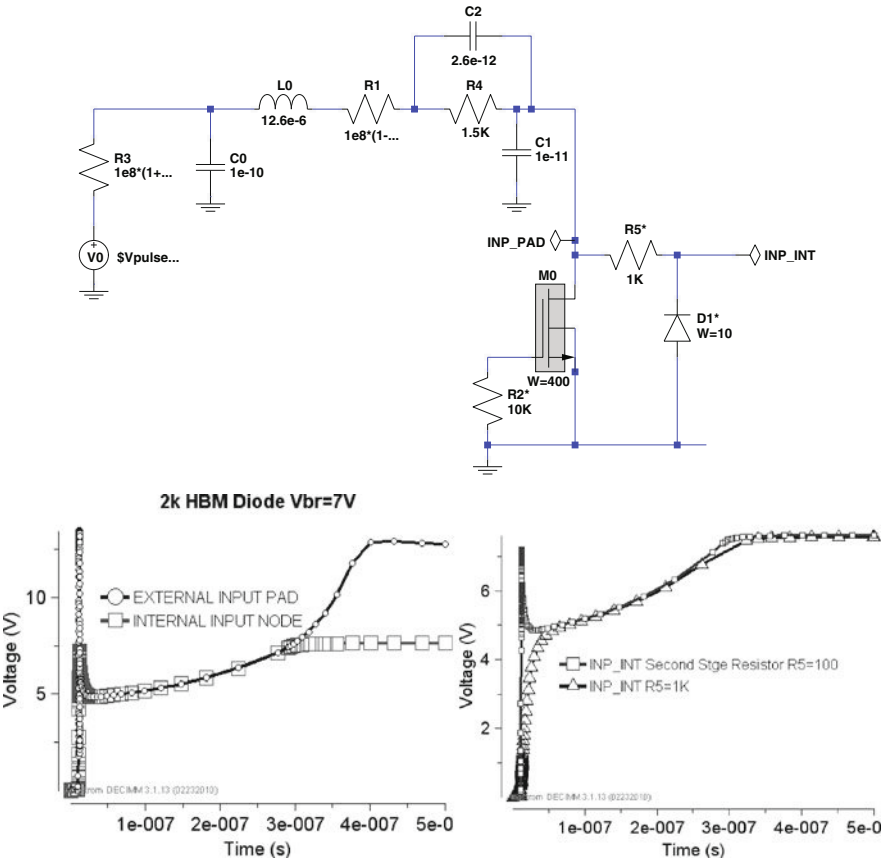


Fig. E4.3 Mixed-mode simulation circuit and comparison of the 2 kV HBM waveforms at the external and internal input nodes

resistor, and 7 V avalanche diode. Waveforms at different circuit nodes are compared for different avalanche diode breakdown voltages and the second stage resistor values (Fig. E4.3), demonstrating reduced overstress of the internal input node that is connected to the avalanche diode. Voltage overshoot reduction is important for CDM ESD pulse events.

Example 4.4 High-Voltage NLDMOS-SCR Clamp with High-Side Avalanche Diode Reference

Library Name: Examples4_Snapback_Clamps
Project Name: E4.4_Snapback_HSCRZ_HBM

In this example, the effect of the high-side avalanche diode reference is demonstrated using default components for 2 kV HBM pulse conditions. The connection of the reference diode is enabled or disabled by changing the *R*5 resistor value.

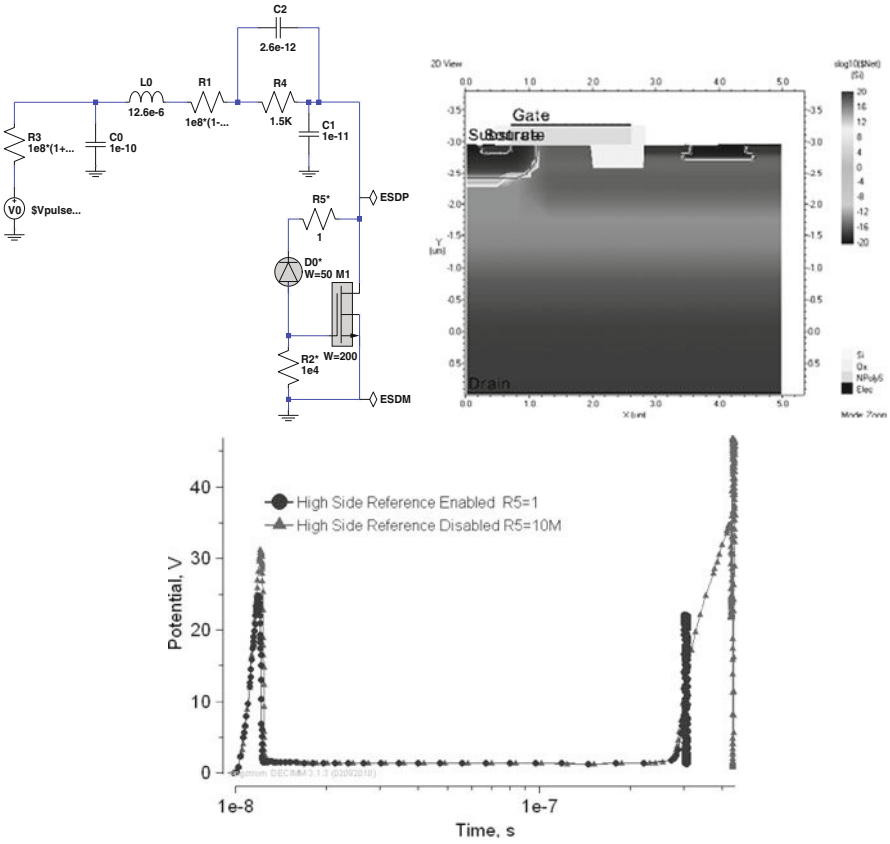


Fig. E4.4 Mixed-mode simulation circuit demonstrating the high-side reference effect in the NLDMOS-SCR clamp and the change in the triggering and turn-off waveforms

Comparison of the HBM pulse waveforms demonstrates the effect of the diode on both the triggering voltage and the residual voltage after turn-off in the case of a high-impedance load (Fig. E4.4). Typical analysis includes optimization of the clamp components for appropriate matching of the avalanche diode current, and gate resistor and snapback device parameters to achieve low d*V*/d*t* triggering effect.

Example 4.5 PNP Clamp with Low Side Avalanche Diode Reference

Library Name: Examples4_Snapback_Clamps
Project Name: E4.5_Snapback_PNP_HBM

In spite of a larger footprint, the PNP clamp is typically preferred for meeting high holding voltage requirements. This example allows comparative analysis of the PNP clamp with other high-voltage device clamps. The clamp is analyzed both with and without the low-side avalanche diode reference under HBM ESD pulse conditions. The low-side reference avalanche diode is enabled and disabled by varying the

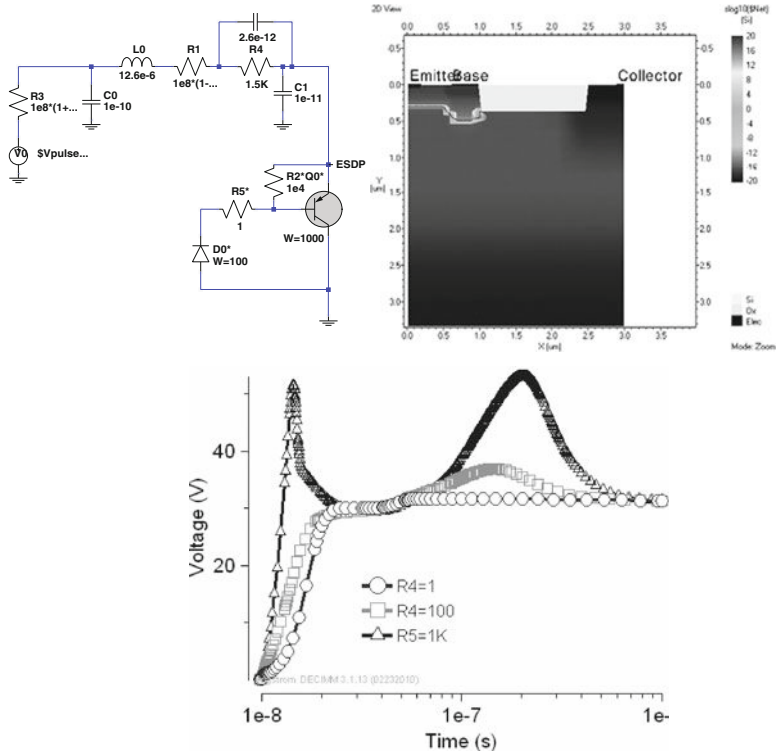


Fig. E4.5 Mixed-mode simulation of the low-side avalanche diode referenced clamp under 2 kV HBM ESD pulse with different values of resistor *R*5. A 30 V compact model is used for the avalanche diode in this example

value of resistor *R*5. Resistor *R*5 also adds on-state resistance to the ideal avalanche breakdown characteristics produced by the compact diode model used for D0 in this example. Comparison of the response to the HBM ESD pulse demonstrates the impact of the low-side reference on both the triggering voltage and residual voltage after turn-off in case of a high-impedance load (Fig. E4.5). Further analysis may include optimization of the clamp components for appropriate matching of the avalanche diode and gate resistor to achieve a low d*V*/d*t* triggering effect.

Example 4.6 High-Voltage NPN Clamp

Library Name: Examples4_Snapback_Clamps
Project Name: E4.6_NPNZ_HBM

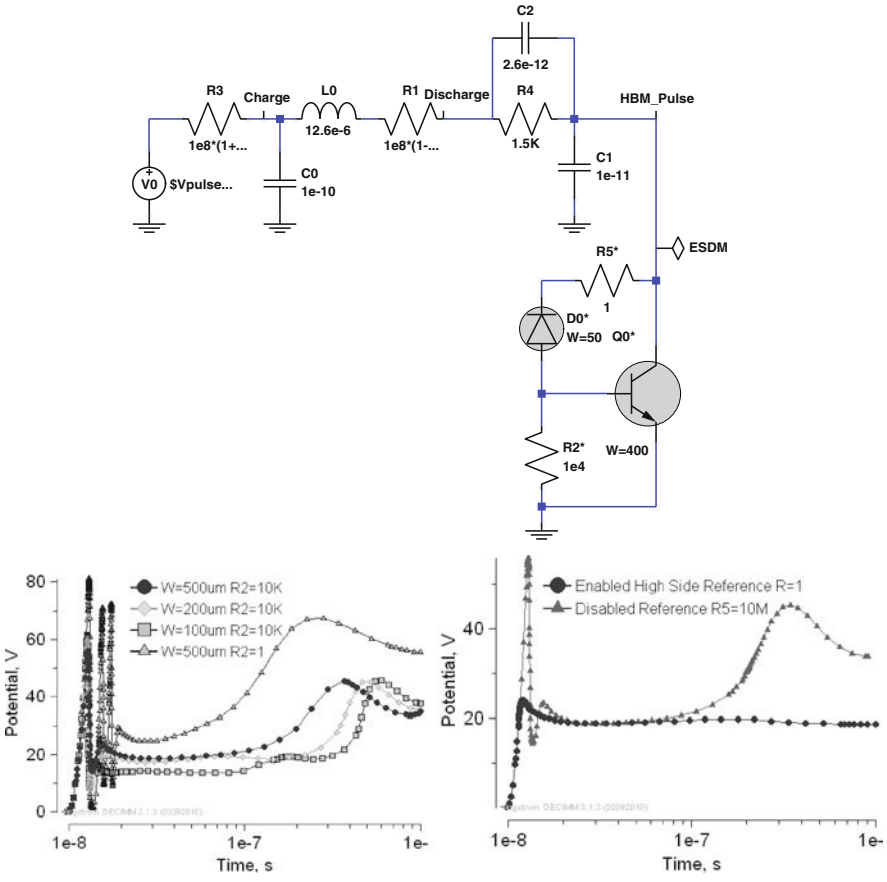


Fig. E4.6 Mixed-mode simulation circuit for the NPN clamp with high-side avalanche reference diode; voltage waveforms produced by the NPN BJT clamp under 2 kV HBM ESD pulse conditions as a function of transistor width and base resistor values illustrating the effect of the high-side diode reference

This simulation example of an ESD NPN BJT clamp studies clamp operation with different high-voltage NPN parameters and clamp component values under a 2 kV HBM ESD pulse. This clamp includes a high-side avalanche diode that can be enabled or disabled by changing the value of resistor *R*5. The impact of circuit component values on both the triggering voltage and the residual voltage after turn-off in case of a high-impedance load is shown in Fig. E4.6. This example can be used for optimization of clamp component values to achieve desired performance.

Example 4.7 Bipolar SCR ESD Clamp

Library Name: Examples4_Snapback_Clamps
Project Name: E4.7_BSCR_HBM

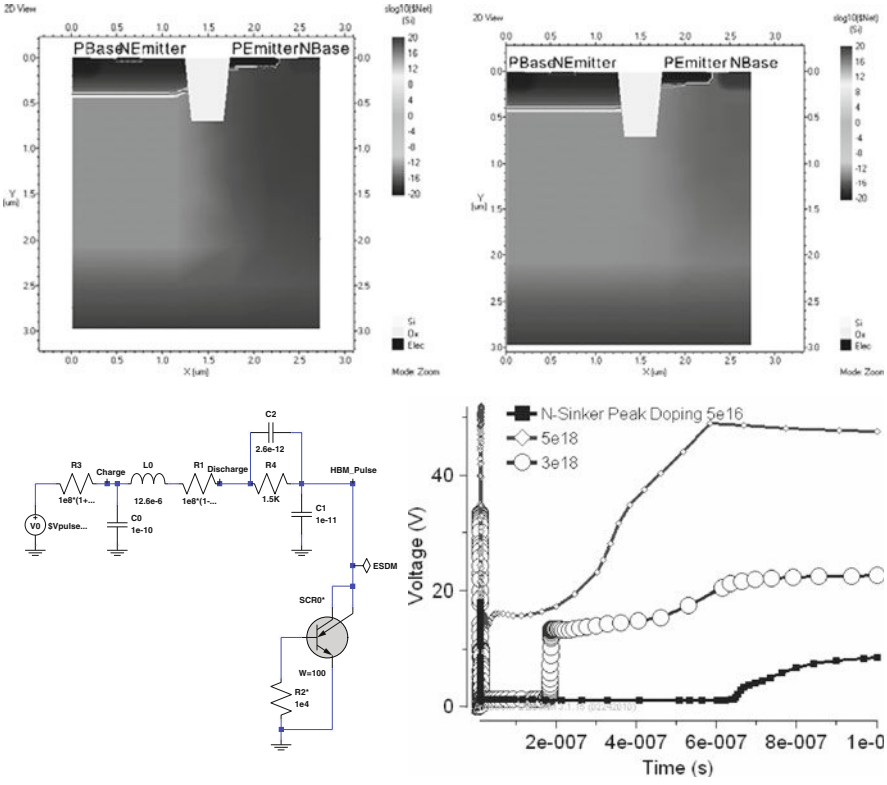


Fig. E4.7 Cross-sections of the bipolar SCR with different N-sinker peak doping values, HBM mixed-mode simulation circuit for the bipolar SCR clamp, and simulated waveforms under the 2 kV HBM ESD pulse for different values of peak N-sinker doping level

This example provides a parameterized clamp circuit for the high-voltage bipolar SCR clamp. The high-voltage bipolar SCR device is obtained from 20 V NPN BJT device by embedding an additional P-emitter region isolated by an N-sinker. The impact on the N-sinker peak doping level on the clamp holding voltage is shown in Fig. E4.7. This example demonstrates the change of both the triggering and turn-off characteristics of the device as a result of different doping levels of the N-type isolation region around the P+-emitter. Alternatively, in practical cases, the N-sinker implant location can be changed to reduce the level of P+-emitter isolation.

Example 4.8 Diode-Triggered SCR ESD Clamp

Library Name: Examples4_Snapback_Clamps
Project Name: E4.8_DTSCR_HBM

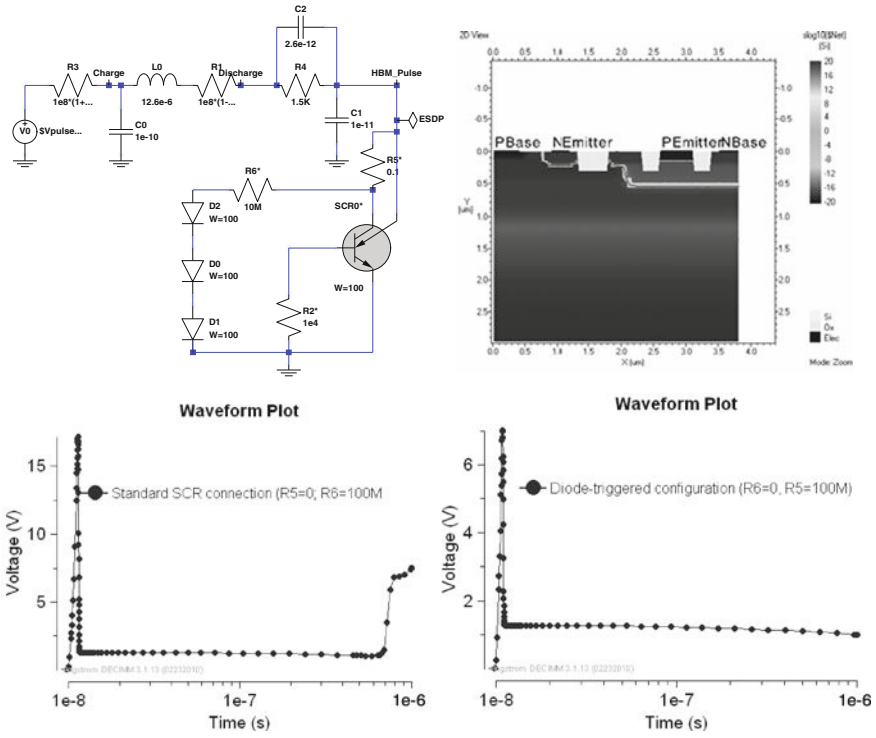


Fig. E4.8 Combined mixed-mode simulation circuit for the standard and diode-triggered SCR, cross-section of the SCR implemented in the CMOS process, and voltage waveforms under 2 kV HBM ESD pulse, demonstrating the capability of the DTSCR clamp to reduce voltage stress

A diode-triggered SCR is a key solution in the case of low-voltage pin protection. It produces low-voltage waveforms to protect the gate oxide in CMOS circuits. The example provides a parameterized SCR clamp circuit. The P+-emitter region of the SCR is directly connected to the protected pin. Different options for connection of the N-base electrode are realized by changing the values of resistors *R*5 and *R*6 to study both standard SCR and DTSCR modes of operation in a single example circuit. The case of a high-side N-base connection is compared to the case of an N-base connection to the ground via stacked diodes (Fig. E4.8).

Chapter 5
ESD Network Design Principles

This chapter covers material needed for understanding the next level of the ESD design hierarchy – the protection network. The protection network or protection circuit is usually composed of ESD protection clamps (cells) connected together in a way that provides a high current path for all of the pin-to-pin combinations. This network is engineered based on certain general principles and assumptions that are discussed below.

In spite of a high variety of analog circuits, these major principles of network design are rather universal. This fact is demonstrated in detail in this chapter, followed by Chapters 6 and 7 that describe ESD protection network implementation in analog signal path and power circuits with the help of practical examples.

The ESD protection network can further be understood as a virtual pulsed power circuit composed of ESD clamps, appropriate high-current-capable metallization routing, and components of the internal circuit of the product that interface with the protected pins.

The main function of the network is to react to a high transient voltage and create a discharge current path. The discharge current path should limit the voltage level in the ESD time domain that suitably guarantees survival of the internal circuit components.

Depending on internal circuit specifications, both transient-triggered and voltage-reference clamps can be used to engineer the ESD protection network. Moreover, in general, some internal circuit components that directly interact with protected pins can also be used to provide a complete or partial discharge current path, thus representing ESD protection network components.

The term *ESD pad ring* is often used in digital circuit design. The ESD pad ring includes pads and a distributed ESD protection circuit created near the pads by a combination of cells. ESD protection network components can be embedded in the I/O cells. In most cases, the ESD pad ring can be decoupled from the internal circuit and is capable of supporting different internal circuits.

For analog circuits, understanding ESD pad ring design is not always as straightforward as for digital circuits. The new level of complexity is the result of increased interaction of the current path with the internal circuit.

In principle, an analog circuit pin can interface with many more connected active devices. Often, deep inside the layout of the internal circuit, some components may

V.A. Vashchenko, A. Shibkov, *ESD Design for Analog Circuits*,
DOI 10.1007/978-1-4419-6565-3_5, © Springer Science+Business Media, LLC 2010

see a significant ESD signal. Such a signal can be transmitted, for example, through the parasitic capacitance of the integrated power devices. This situation may require additional protection measures for the internal circuit nodes. Alternatively, in the case of micro-SMD (surface-mount device) design, bump pads can be scattered on top of the whole active layout area. Thus, the definition of the ESD pad ring for analog design is rather "fuzzy."

Nevertheless, although ESD network generally may not be decoupled from the internal analog circuit, it can still be analyzed to a certain level using hierarchical methods.

ESD pulse voltage limiting is targeted below the absolute maximum pulsed voltage rating of the devices interfacing with the pins. At the same time, during ESD stress, the transient conditions of the control electrodes and internal circuit nodes are generally unknown.

This conclusion is quite important. In Chapter 3, it has been demonstrated that pulsed SOA of most standard active devices has a critical burnout voltage that is dependent upon control electrode conditions. For example, in N-channel CMOS devices, due to the multiplication of the channel current, the critical voltage for an irreversible snapback event has a strong dependence upon the gate bias. This voltage is much higher below threshold, in comparison with open channel conditions. Similar dependence is observed in NPN BJT, here as a function of the base current (Chapter 3). Therefore, depending on the internal circuit design, the absolute maximum voltage of the internal components may vary significantly, impacting the ESD protection window voltage range or even making it negative. Specifically, this aspect is mainly responsible for why analog circuits often require custom ESD protection networks or even custom clamp design.

Unfortunately, the conditions of the internal circuit are hard to predict or simulate. Accuracy or even the possibility of ESD pulse simulations is usually limited by the absence of ESD snapback compact models, while an empirical analysis is complicated due to the many floating points in the circuit.

The most robust solution for a simplified case is presented by a scenario in which the voltage drop provided by the ESD protection network at the pin is below the absolute maximum pulsed voltage for all the connected components at any allowed control electrode conditions. This case is typical for high precision amplifiers, for example, where the requirements for linearity of the components often automatically provide a significant range between operational maximum and absolute maximum voltages.

However, in the case of power-optimized analog circuits, the opposite situation may arise. For example, in fast switching pins, due to the "aggressive" SOA of switching power devices, the critical voltage in the on-state might be even lower than absolute maximum voltage specified for the circuit.

As it will be demonstrated (this chapter and Chapter 6), the ESD protection network is specific to its analog circuit. This implies a necessity of new experimental evaluation of the product's ESD performance each time the internal analog circuit is modified. If verification results are negative and exhibit underpass issues, the next

challenge is in detecting the reason for the issues and proposing and validating measures or solutions that would fix the problem. In this case, often several alternative solutions can be generated, each with a different level of confidence.

The ability to make the right choice greatly depends upon the understanding of ESD protection network design and the components used. This is a practical alternative to the comparison of new results with previous case studies that may not be equivalent. Choosing the right corrective option from the several proposed ones especially becomes a challenge in the case of product-dedicated tapeout. The mask set may be rather expensive or a full change of the mask set may not fit the business timeline. Often, running multiple die reticle (MDR) to validate different experimental ESD versions may not be a practical option. Under such circumstances, it is often desirable to fix the problem using a minimal number of masks or by utilizing metal options that can be applied to the wafer preserved before the backend process steps.

According to the authors' experience in the field, ~80% of all underpass issues are related to mistakes that could have been avoided if simple principles and known cases were studied and were taken into account.

The purpose of this and the following chapters is to provide the readers with corresponding guidelines that will help them in practical design. One of the most powerful tools in validating both ESD networks components and fixing options is the mixed-mode numerical simulation. Several simulation examples for an ESD protection network using DECIMM^{TM} complement this chapter.

5.1 Rail-Based ESD Protection Network

5.1.1 Rail Based and Local ESD Protection

In order for an IC chip to pass the ESD qualification for the typical package level specification (e.g., 2 kV HBM, 200 V MM, 750 V CDM), each individual pin must be connected to the ESD protection network. In most cases, the type of application determines the choice of ESD protection.

For a single voltage domain, the design of ESD protection networks can be roughly subdivided into *rail-based* and *local* ESD protection types. The advantages and drawbacks of both approaches are summarized in Table 5.1.

Local ESD protection is independent of bus resistance, as well as that of other network components. It provides an easily predictable local voltage waveform at every pin. However, especially in the case of high voltage protection or, alternatively, in the case of high pin count, local clamps may not be the most optimal space-saving solution. In addition, local clamps are generally more process-sensitive or may greatly interfere with normal circuit operation.

A rail-based ESD protection network can be composed of a *snapback* power clamp or a distributed network of *active* clamps and ESD diodes. Here, *snapback* clamps are contrasted with *active* clamps (Section 4.1) to emphasize the

Table 5.1 ESD protection strategies

	Local snapback	Diodes and active clamp
Main advantages	Independent of bus resistance Local to every pin	Area efficient Highly process portable SPICE-based design
Main drawbacks	More area intensive Highly process sensitive Difficult to simulate	Performance depends on bus resistance Distributed ESD network requires complex chip level implementation and verification

corresponding difference between the uncontrolled current in the conductivity modulation mode and the controlled current in the active operation mode of standard devices.

In the case of distributed rail-based protection with active clamps and ESD diodes, area efficiency depends on the pin count. For a product with a high pin count area efficiency can be rather high, while for a product with fewer pins the solution can be absolutely inadequate, even in the same voltage domain.

In principle, local protection can also be realized with active clamps, but this solution is impractical.

The major advantage of the active clamp network is the possibility to use circuit simulation to analyze performance under ESD pulse conditions. At the same time, at a certain level of complexity, verification of the distributed ESD protection network becomes a challenge.

Rail-based protection with a snapback power clamp is useful for high-voltage analog circuits, for example, amplifiers. If the circuit is low voltage, then the rail-based ESD protection network can be arranged using active clamps as described in Section 4.1. The active NMOS clamp provides a significantly lower clamping voltage due to its design principles.

ESD protection can be embedded in the input/output (I/O) cells. Under this approach, a set of ESD and latch-up rules must be taken into account to provide a non-conflicting operation of ESD elements and I/O cell components.

While the ESD compliance of each unique I/O cell is simply ensured by design, ESD robustness of each product on the chip level can be achieved by the following set of ESD-related rules during I/O ring integration. In the past, compliance of the I/O ring with all the ESD rules had to be verified in a time-consuming and error-prone manual review process.

To facilitate this process in high pin count digital domains, an automated tool is usually developed that provides an automated assembly and verification of the I/O ring for compliance with all ESD-related integration rules, in both wire-bond and flip-chip designs.

In the case of rail-based network, there are several steps in the complex chip verification. They usually include verification that every IC pin includes ESD

protection; analysis of the dedicated ESD discharge paths; extraction and verification of interconnect robustness (contacts, metal, and vias) for ESD current density. The verification can also be completed for ESD cells to confirm that the cells are properly sized and their placement is correct.

There is a number of cases where active clamps cannot be used due to the fast transient conditions at the power supply pin. Another conflicting instance that limits active clamp use is in high-voltage circuits, where the size of the adequate distributed network components is too spatially inefficient.

5.1.2 Rail-Based ESD Protection Using Snapback Clamps

Any rail-based ESD network can be identified by the presence of ESD busses (rails) with a common clamp and the ESD diodes connecting the other pins to the rails (Fig. 5.1). ESD diodes can connect input/output (I/O) pins or other control pins, including power supply pins. The power supply pins connected to the rails typically share the metallization routing. However, in principle, the network can be organized with ESDPLUS and ESDMINUS busses that separate the power supply and ground for the domain.

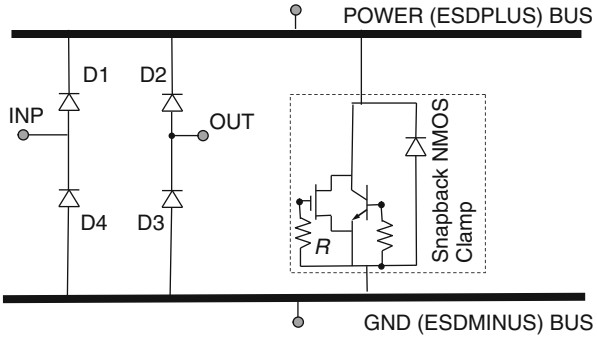


Fig. 5.1 Rail-based ESD protection network

In rail-based ESD protection, the ESD diodes are designed to conduct pulsed ESD current only in the forward bias mode. To avoid interference with the major ESD current path, the voltage tolerance of the ESD diodes is an important parameter that needs to be met. An appropriate breakdown voltage of the rail-based protected domain – above the power supply level – should be provided to avoid ESD diode burnout due to avalanche breakdown mode and to minimize direct impact on circuit functionality in normal operation regime. For example, the p^+ -Nwell diode with a breakdown voltage of ∼11 V should not be used to protect a 12–10 V erase pin EEPROM circuit block.

An example of the simplified four-pin rail-based circuit is illustrated in Fig. 5.1. The network provides protection of the input (INP), output (OUT), POWER, and ground (GND) pins. This network can represent an amplifier. Additional pins can be added to the network by reusing similar diode connections to the I/O pins, as shown in Fig. 5.1. However, the higher pin count will not change the analysis below.

For each pin-to-pin combination, this rail-based network forms an ESD current path through the forward-biased ESD diodes, corresponding parts of the rails, and the power clamp. The power clamp provides both the high-voltage current path and the reverse path through the clamp diode. For example, in the case of positive zap of INP vs. OUT, the current path is formed through the upper diode D1, the upper part of ESDPLUS bus, then through the snapback clamp, the corresponding part of ESDMINUS bus, and the forward-biased diode D4 (Fig. 5.1).

Rail resistance should also be taken into account at ESD current level; the total voltage drop on the circuit and the voltage drop at each pin relative to the ground can be estimated. Results of the calculation are summarized in Table 5.1. These results are obtained under the assumption that the clamp voltage drop is equal to V_C and the voltage of the I/O and clamp forward-biased diodes is equal to V_D, with the contribution of bus resistances $R_{ESDPLUS}$ and $R_{ESDMINUS}$ into the current path.

The table below can be extended for a more complex case or modified for additional accuracy. Higher accuracy can be achieved by taking into account the distributed bus resistance extracted by CAD tolls from exact metal routing.

Similar reference tables are often used for ESD checkers implemented in the CAD environment. The checkers also extract the parasitic resistance and capacitance of the interconnects [104].

Several important conclusions can already be made through the analysis presented in Table 5.2.

Table 5.2 Rail-based peak voltage at peak ESD current

ESD zap combination	Total voltage drop on zap pins	INP voltage vs. ESDMINUS	OUT voltage vs. ESDPLUS	POWER voltage vs. ESDMINUS
INP to ESDPLUS	V_D	V_D	0	0
PWR to ESDMINUS	V_D	$\sim V_D/2$	$\sim V_D/2$	V_D
ESDMINUS to INP	V_D	V_D	0	0
PWR to ESDMINUS	V_C	$V_C/2$	$V_C/2$	V_C
INP to OUT	$2V_D+I_H(R_{E-}+R_{E+})+V_C$	$V_D+I_H(R_{E-}+R_{E+})+V_C$	V_D	$\sim V_D$
INP to ESDMINUS	$V_D+I_H(R_{E-}+R_{E+})+V_C$	$V_D+I_H(R_{E-}+R_{E+})+V_C$	$(I_H(R_{E-}+R_{E+})+V_C)/2$	$I_H(R_{E-}+R_{E+})+V_C$

First of all, the biggest overstress occurs for the INP–OUT pin combination. Therefore, the diode and power clamp components at that combination should provide a low on-state resistance to limit input voltage overstress at ESD pulse. In this

case, the major components are the I/O diodes, which introduce signal limitation due to parasitic capacitance.

Another important consideration that cannot simply be neglected is the ESD bus resistance. Usually, resistance of first metal layers is ~30 mΩ per square of the metal area. For chip die dimensions of 4 mm × 4 mm, the total bus length at the corner opposite to the clamp will be ~8 mm. In the case of a closed ESD pad ring, the bus is routed along both sides of the chip and should be accounted for twice. If a single metal bus with a 100 μm width is used (or an equivalent double metal layer bus of a 50 μm width), the total accumulated resistance in the current path will be 1.2 Ω. Thus, at a 2 kV HBM 1.33 A ESD current, the protected pin opposite to the power clamp in the chip will accumulate an additional voltage drop of ~1.5 V. In principle, this level is acceptable for a 5 V process, but might create a problem in the 1 V process. If the bus width is only 10 μm, the corresponding 10-fold higher voltage drop can result in immediate ESD failure of protected pins opposite to the power clamp in the die.

One of the major disadvantages of rail-based protection is the cross-talk between the input and output, mainly through the parasitic capacitance of ESD diodes.

Rail-based protection with a snapback power clamp has already been shown in Fig. 5.1. In opposite to the circuit with an active clamp component, this configuration relies on a parasitic device in high-current conductivity modulation mode.

As was discussed in Chapter 3, in a 0.5 μm CMOS process, for example, the lateral parasitic n–p–n structure of the NMOS can provide a current of 3–5 mA/μm. Thus, the total width of NMOS array required for a 2 A ESD current is ~400 μm.

Unlike the active clamp, the snapback device is rather sensitive to layout and process variation. At the same time, the voltage of the clamp cannot be reduced below the holding voltage of the NPN, thus providing a limiting factor of the network clamping voltage.

Rail-based protection with a snapback power clamp is especially efficient in bipolar processes, high-voltage products, transient power supply, or voltage domain operation, and in the case of low pin count. It is advantageous in case of critical requirements for the silicon space, when active clamp rail-based protection becomes no longer optimal.

5.1.3 Rail-Based ESD Protection Using Active Clamps

Active clamps rely on the device working in normal operation mode [60–70]. The original and most typical implementation (Section 4.1) of this clamp is in an oversized low-voltage NMOS array capable, in the on-state, of safely sinking the current generated by HBM, MM, or CDM pulses for typical precharge voltages [60].

In the case of a 5 V 0.5 μm process, the typical ratio of the total NMOS width and length, *W*/*L*, is 5000/0.5. This ratio is required to provide a saturation current of ~2 A. Bus resistance plays an important role and must also be included in simulations. The RC-timer time constant typically selected is ~6 μs, which keeps the

active clamp in the off-state during normal operation and, at the same time, avoids significant overload during voltage increase at the powering sequence.

Both the distributed NMOS array and RC timer are space-consuming clamp components (Chapter 4). The clamping voltage depends on the on-state resistance of the clamp and directly on NMOS width. For the 5 V process, active clamps are usually designed with an NMOS voltage drop below ~2 V. This provides an appropriate design margin for the voltage drop on the I/O and clamp ESD diodes and metal busses. For extremely low parasitic capacitance at the I/O pin the rail clamp can be oversized in order to provide a much lower clamping voltage, while the size of the ESD diode can be significantly reduced. In this case the additional voltage drop on the diode will be compensated by reduction of the clamp voltage, providing an advantage in lower equivalent capacitance at the protected pin.

The most typical applications for rail-based ESD protection with an active clamp are low-voltage circuits. These circuits provide CMOS components with an appropriate voltage tolerance to the power supply level, a conventional (relatively slow) powering sequence for the power supply and input pins, and a sufficient pin count. In this case, the advantageous low on-state resistance allows the designer to build a distributed network of RC-controlled power NMOS devices with a total width of ~3000–5000 μm.

A simplified schematic and layout of the rail-based active clamp ESD protection network is presented in Fig. 5.2. A more detailed description of active clamp design principles and operation is given in Section 4.1. It is practical to compose the active clamp library of standard cells that include "slave" active clamp cells and an RC-timer cell. In the case of high pin count, the RC-timer cell can be placed in the corner

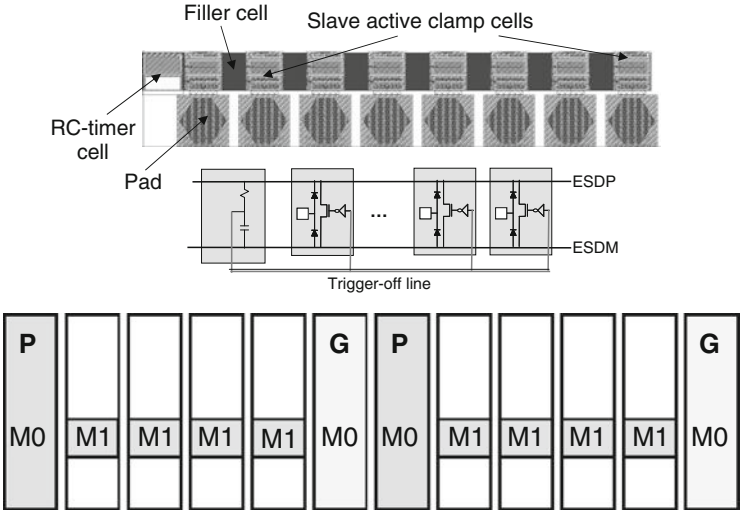


Fig. 5.2 Example of layout and schematic views for 3 V ESD domain obtained with redesign according to new design rules in active clamp ESD library [104]

of the chip. Compatible filler cells can provide the connection between the "slave" active clamp cells and the RC-timer cell. The filler cell continues the bus connection for ESDPLUS, ESDMINUS, and the trigger line (Chapter 4) and can accommodate an additional NMOS device.

Due to the large size of the NMOS device, the active clamp is usually designed as a distributed network across the pad ring area, occupying $\sim 100 \times 100$ μm at each pad. At the typical pad dimension of 70–120 μm, each "slave" active clamp cell usually accommodates a 500–1000 μm wide NMOS with a turn-on capacitor of ~500 fF and a pair of ESD diodes. An optional trigger line capacitor can be added to compensate for the voltage drop across the trigger line. In this case, at least 8–10 clamps are necessary to provide appropriate voltage clamping in a single voltage domain across the ESD ring.

An example of part of the pad ring schematic is shown in Fig. 5.3. Each I/O PAD has a diode connected to ESDPLUS and ESDMINUS. An RC trigger cell acts as a low-pass filter and turns off the active clamp during normal operation. The RC time constant is set to 6 μs. When the chip is powered up, with a power supply voltage-ramping time constant above 6 μs, ESDPLUS gains a higher potential relative to ESDMINUS and the V_{TRIG} node also remains at high potential.

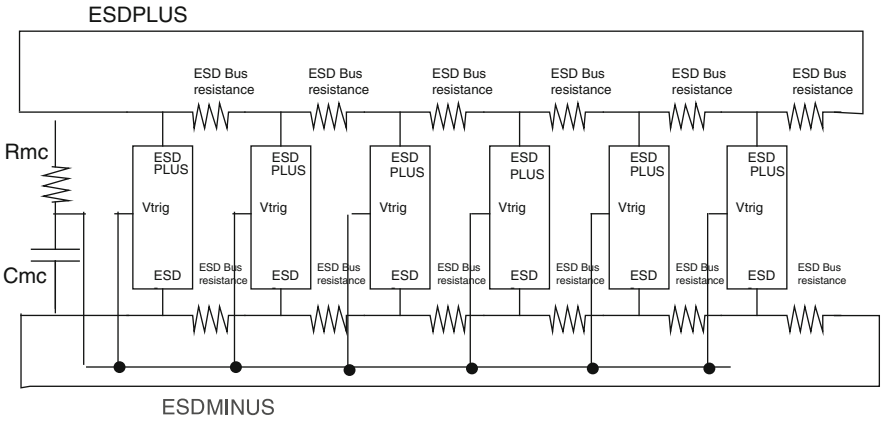


Fig. 5.3 Distributed Merrill clamp scheme suitable for 5 V ESD protection

The active clamp protection scheme is advantageous because the clamp voltage is proportional to the number of distributed cells used in the design. If the clamp voltage is below VDD, then there is no requirement for current-limiting resistors at the input or output, or indeed any other precautions. The disadvantage of the scheme is the large area required for the protection network.

Clamp usage is limited if chip specifications require a "hot plug-in" compatibility. The possible fast rise time during hot plug-in events may activate the ESD protection, resulting in voltage overstress of the clamp. In this case, the snapback solution is required.

I/O ESD compliance of individual I/O cells is achieved by the following ESD integration rules.

In practical application, the active ESD clamp solution is often designed in conjunction with I/O library cells, as this is an efficient use of chip space (Fig. 5.4). I/O circuits in analog IC typically contain many different I/O cell types that are optimized for the many different types of pins used in the circuit, for example, high-speed, power, high-voltage, EEPROM, digital, LVDS, and USB pins.

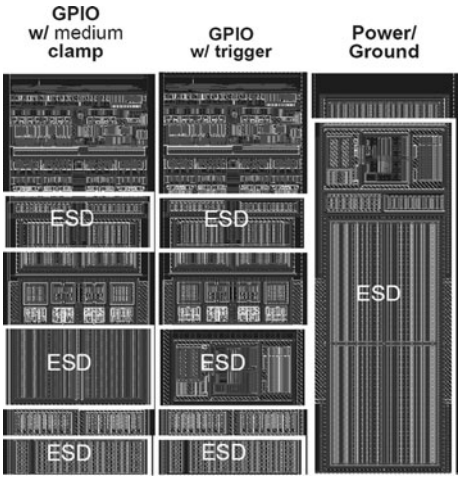


Fig. 5.4 Example of I/O and active clamp layout with mark-up of the area used for distributed ESD protection [104]

The complex design of the distributed pad ring requires verification of many sets of ESD integration rules. Due to a high variety of I/O multiple verifications of the whole pad ring are required. This is often achieved through automated extraction and checker software [104] implemented in the CAD environment.

In a properly built active clamp protection with appropriate bus resistance, the ESD-current-conducting NMOS operates in normal operation mode, far from avalanche breakdown conditions. Therefore, at overload, there are usually no failures observed in this component. Perhaps one of the most vulnerable active clamp components is the speed-up capacitor that is responsible for NMOS turn-on during the fast rise time of the ESD signal. Therefore, a robust operation is required from this component. For example, the most process-sensitive parts of the capacitors are corner regions. An example of ESD damage location is presented in Fig. 5.5.

Other process-related issues are associated with the large NMOS itself. Often, in small analog chips, the Active Clamp NMOS is the device with the greatest total width, while the internal circuit may be either mainly BJT-based or include a negligible amount of the minimum-width input NMOS devices. In this case, the yield of the product may be significantly impacted by the choice of ESD solution, due to the different yield of the large NMOS array in comparison with that of relatively small snapback NMOS clamp.

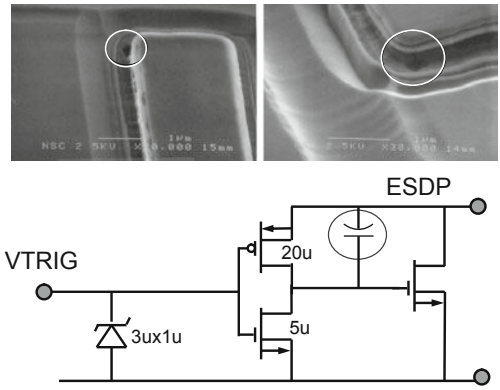


Fig. 5.5 One of the "weak spot" failures in the active clamp as a result of irreversible speed-up capacitor breakdown. *Circle* indicates the capacitor which had a leakage between the P^+ polytop plate and P-sinker bottom plate

5.1.4 Specific of Active Clamp Design in BiCMOS Processes

5.1.4.1 Verification by Circuit Simulation

A circuit simulation with extracted parasitic metallization resistance and capacitance is required to ensure that the clamping voltage is sufficiently low. In addition to the simulation verifying the signal in normal operation regime, three types of ESD simulations are usually required.

The first HBM source simulation targets the single pulse waveform to verify the low clamping voltage level between ESDPLUS and ESDMINUS. The second simulation is done for the sequence of ESD pulses between I and O to verify that the clamp does not suffer from a voltage accumulation effect. This effect can be observed irrespectively of whether the tester provides a short circuit condition between zaps, since the I/O diodes separate the clamp from the tester. This simulation is important because it ensures that the clamp keeps the voltage low for chips with a high pin count when multiple ESD zaps are done during ESD testing.

Finally, a simulation should be performed for the power supply sequence. This simulation is done to ensure that the clamp does not drain an excessive current during the expected ramp of the power supply and provides low leakage after achieving the power supply voltage level.

Most popular design solutions are based on the NMOS acting as a clamping device. The circuit for this solution has been discussed in Section 4.1.

To illustrate circuit simulation aspects in this section, more examples for active clamps are analyzed based upon NPN BJT and PMOS devices for a 3 V BiCMOS process.

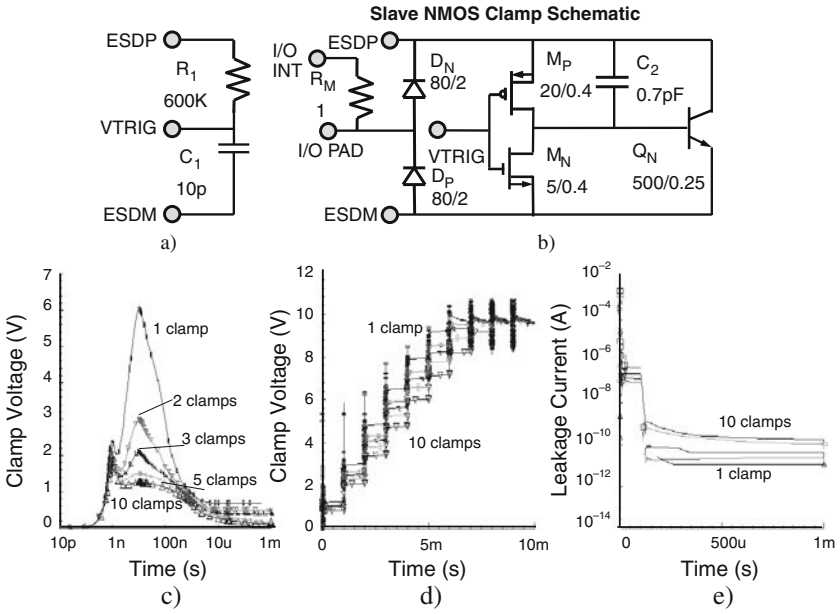


Fig. 5.6 Circuit diagram for the RC timer (**a**) and slave clamp (**b**) cell based upon NPN BJT and the results of a circuit simulation for the single pulse waveform (**c**), multiple zap sequence (**d**), and power supply sequence (**e**)

The NPN BJT clamp has a 3 V CMOS driver with an NPN base directly connected to the inverter output (Fig. 5.6a, b). Each clamp has an NPN with a total emitter width of 540 μm. The RC-timer time constant is ~6 μs.

The clamp provides an appropriate low leakage in the powering sequence (Fig. 5.6e) and the desired single pulse waveform (Fig. 5.6c) with a low clamping voltage, if the number of slave clamps is sufficiently high (Fig. 5.6c). The peak amplitude of the waveform produced by the circuit is below 1 V and practically does not change starting with ~10 clamps (Fig. 5.6c). However, the voltage accumulation effect is observed in the case of a multiple zap sequence through the I/O diodes (Fig. 5.6d).

A similar phenomenon is observed in the 3 V PMOS-base active clamp (Fig. 5.7), unless the clamp count is significantly increased. The clamp uses the same RC timer (Fig. 5.6a), and the slave clamp is built upon the PMOS with a total width of 1368 μm (Fig. 5.7a). With the increase of the clamp count, the problem of the accumulation effect is solved (Fig. 5.7c). The accumulation effect physically represents the effect of incomplete discharge between the rails. The effect can be significantly different, depending on the power consumption and leakage of all the different circuit blocks that are not included in this section. If the circuit can tolerate the leakage of a simple "bleed" resistor connected between the rails, the accumulation effect is eliminated. Alternatively, a more sophisticated discharge, at low current levels and with a larger time constant, can be proposed.

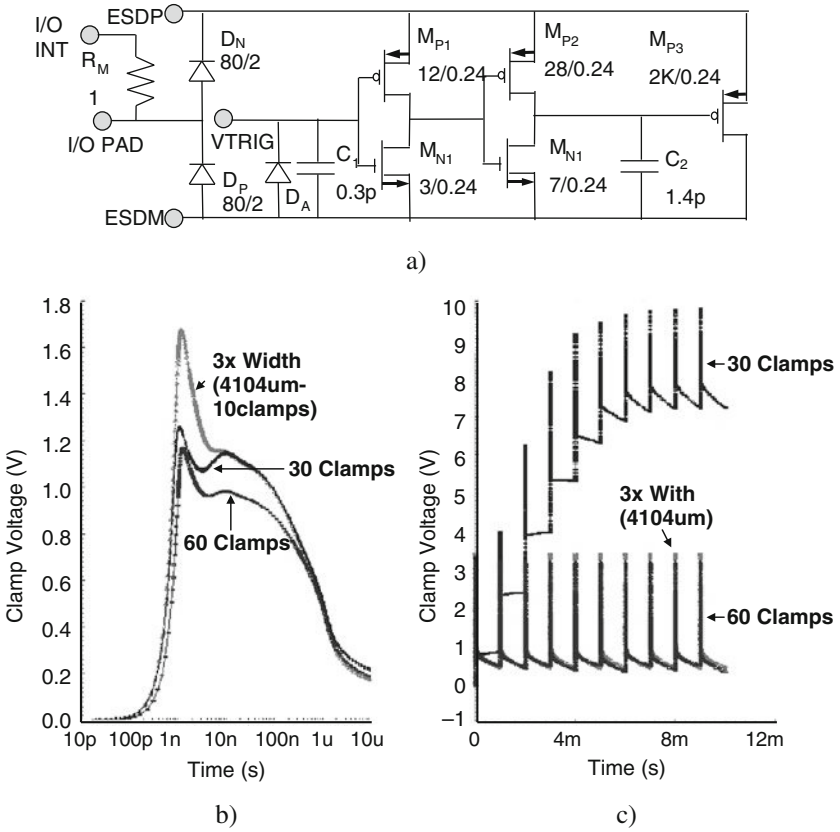


Fig. 5.7 (**a**) Circuit diagram for a slave clamp based upon 3 V PMOS and the results of circuit simulation for the single pulse waveform (**b**) and multiple zap sequence (**c**), depending on the number of clamps

5.1.4.2 Experimental Comparison

In the experimental setup, tests similar to the above simulation analysis are rather difficult to reproduce. The adequate comparison requires access to a waveform capture tester. An alternative way to compare at least high-current performance is by using TLP characterization.

Using standard TLP pulse tools, the ESDP to ESDM combination is compared below for several alternative clamp designs in the BiCMOS process. With appropriate clamp design, the corresponding IO-to-IO or other zap combinations are usually different from the ESDP–ESDM with an additional voltage drop on the two corresponding ESD diodes. A comparison of these characteristics for the case of 10 slave clamps with a total width $W = 540$ μm in a 2.5 V NMOS device is presented in Fig. 5.8.

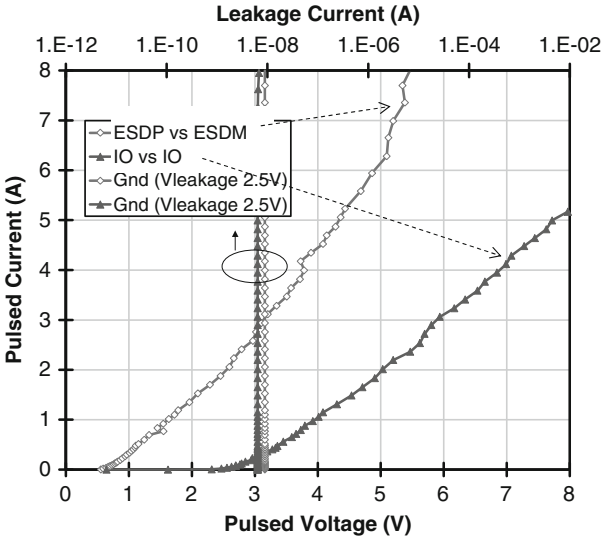


Fig. 5.8 Comparison of the 100 ns TLP characteristics for a 2.5 NMOS-based active clamp in the case of ESDM–ESDP vs. IO–IO zap combinations

An additional important conclusion drawn from the data (Fig. 5.8) is that the limitation of the clamp is not related to the level of the critical current I_{T2}. A network assembled from 10 clamps can easily provide a current level that is significantly higher than the one dictated by the corresponding standard package-level specifications requirements, for example, in the 2 kV HBM pulse. This level corresponds to a peak current of ~1.33 A, which creates a voltage drop of ~4.5 V in the example of IO to IO (Fig. 5.8). On the contrary, the limitation of the clamp is related to the voltage waveform produced by the clamp. An appropriate low voltage is achieved by increasing the number of clamps in the network.

The effect of the linear width scaling for the clamp design previously analyzed by the numerical simulation (Fig. 5.6) is demonstrated in Fig. 5.9.

As expected, a better high-current capability is observed for the NMOS-based active clamp vs. the PMOS-based active clamp, due to the better high-current characteristics of the n-channel device (Fig. 5.10). Another major conclusion supported by the experimental data is the comparably enhanced performance of NMOS-driven BJT-based slave clamps with respect to pure CMOS active clamp design, as demonstrated in Fig. 5.10.

In the case of non-stacked low-voltage solutions, the BJT clamps provide an advantage mainly due to the smaller area of the BJT device as compared with the same degree of performance. However, in the case of higher voltage tolerance protection, the BJT design provides a significantly superior performance when stacked NMOS clamp drivers are used. This case is described in the following section.

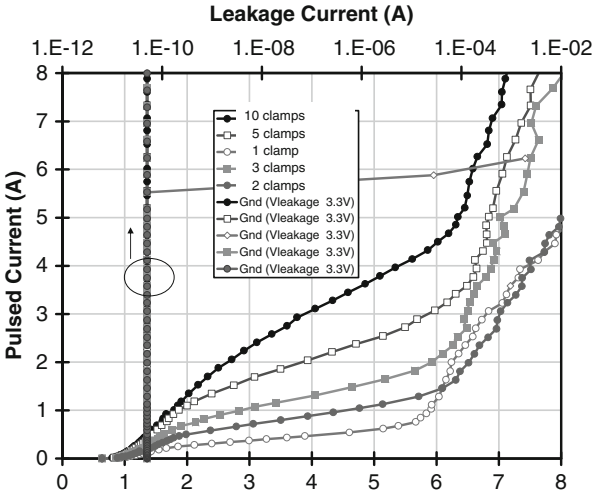


Fig. 5.9 Comparison of the ESDP–ESDM TLP characteristics for an ESD protection network assembled with 10, 5, 3, 2, and 1 slave clamps

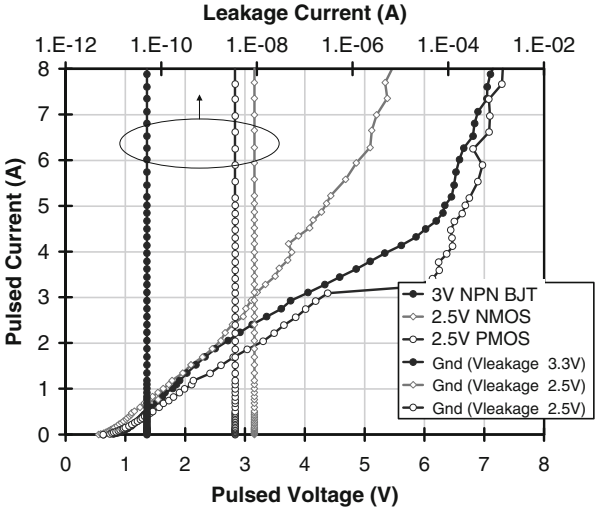


Fig. 5.10 Comparison of the ESDP–ESDM TLP characteristics for the 10-clamp network assembled with 3 V BJT NPN, 2.5 V NMOS, and 2.5 V PMOS slave clamps

5.1.4.3 Active Clamp Protection in Complementary BiCMOS with Low-Voltage CMOS Components

One of the most complex active clamp solutions is realized in the complementary BiCMOS process with mixed voltage tolerance of the CMOS and BJT components.

Similarly, in the case of modular process approach, the designer may select cost-minimizing process options such that the mask set supports high-voltage bipolar devices, but only low-voltage CMOS. In this case, a desirable solution needs to be engineered based upon a relatively large footprint-stacked CMOS RC-timer block and driver, while the active high-current clamp component can be implemented using a BJT device.

A particular example of the solution is presented below for the complementary BiCMOS process with 5 V high-gain SiGe NPN and PNP devices, but with only 2.5 or 3 V CMOS components available. The target requirement for the active clamp solution is to provide an active power clamp implementation for the 5 V power supply.

The starting point for design comparison is the alternative fully CMOS configuration. Such a clamp can be engineered using stacked CMOS devices (Fig. 5.11a). The clamp contains an RC timer based upon a stacked poly-capacitor. The RC timer generates two levels of triggering signals V_{TH} and V_{TL} to control the slave clamp driver. These signals are connected to the two distributed trigger lines across the slave clamp network. Each slave clamp is composed from stacked inverter and stacked power 3 V NMOS devices MNL and NMH (Fig. 5.11a). The operation principle of the stacked clamp is similar to the non-stacked CMOS clamp, discussed above. The speed-up capacitors provide for a rapid turn-on of the slave clamp NMOS devices MNL and NMH when the fast rise time of the ESD pulse voltage is applied between the ESDP and ESDM busses. After the delay time generated by RC timer, the two trigger line signals provide for MNH and MNL turn-off into the high impedance state. At the slow rise time power supply ramp, the RC timer keeps the stacked power NMOS device in the off-state.

The aggressive area-efficient solution for the power part of the clamp is presented in Fig. 5.11. It can be realized using either NPN or PNP devices. A version of the clamp with NPN BJT (Fig. 5.11b) is compared to the clamp version containing PNP BJT (Fig. 5.11c). For practicality, the stacked NMOS clamp (Fig. 5.11a) is reused to provide the low-side driver for the common emitter NPN device (Fig. 5.11b) or the high-side driver for the PNP device (Fig. 5.11c).

The results of the simulations performed according to the methodology described in the previous section and the layout implementation are summarized in Figs. 5.12 and 5.13, respectively. For the given timer and driver designs, both clamp versions demonstrate appropriate voltage waveforms for the 5 V voltage domain (Figs. 5.12c and 5.13c), though better clamping characteristics are realized in the PNP-based solution. A voltage accumulation effect is observed in the network with a small NPN-clamp count (Fig. 5.12d).

The conclusion of the superior efficiency of the BJT-based overvoltage clamp is summarized in Fig. 5.14. A network with 10 stacked NMOS slave clamps is compared to a network with a single 1080 μm PNP BJT clamp designed according to the schematic of Fig. 5.13.

A more detailed comparison of BJT-based clamps is presented in Fig. 5.15 for solution footprint optimization. Six different versions of the clamp, “G,” “H,” “I,” “K,” “L,” and “M,” are cross-compared. In this particular circuit design, the

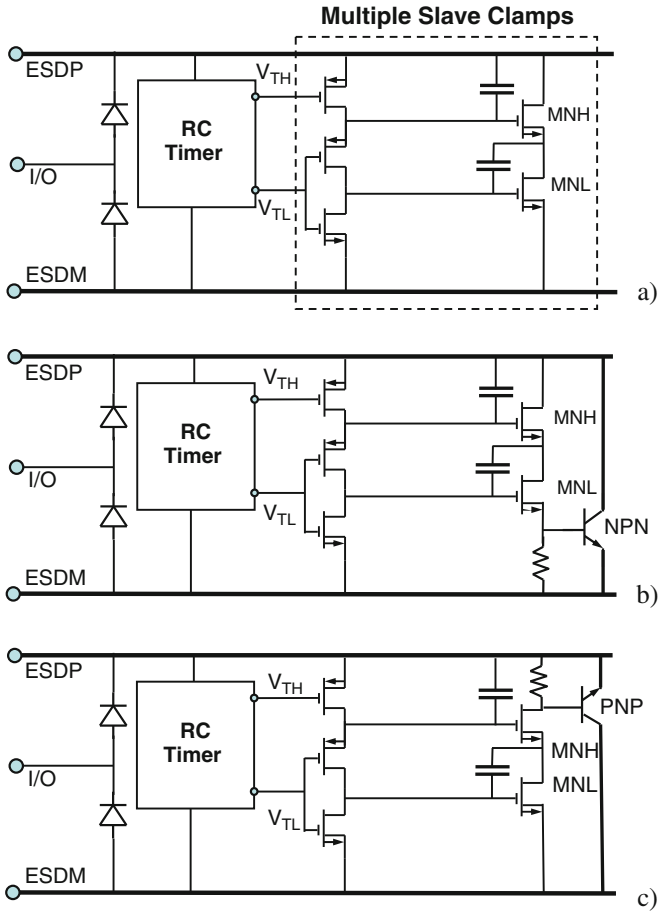


Fig. 5.11 Examples of 5 V active clamp implementation in 5 V complementary BiCMOS process with 2.5 or 3 V CMOS module. Cascaded NMOS active clamp (**a**), NPN based (**b**), and PNP based (**c**)

PNP-based clamps ("K," "L," and "M") provide a better high-current performance in comparison with NPN devices (G," "H," and "I") of the same size, as expected from the above simulation analysis. The lower performance of NPN-based slave clamps in comparison with PNP-based clamps is most likely related to the driver's capability to provide a side base current, since the gain of the NPN devices is higher. Non-linear width scaling is observed in the analysis. A twofold size reduction of the BJT results in only a 30% loss in the on-state current (compare the "H" and "L" clamp versions with the original-sized clamps "G" and "K," respectively). Another way to reduce the footprint of the power clamp is by reducing the size of the driver. However, this measure results in a significant drop of high-current clamp performance (Fig. 5.15, clamps "I" and "M") for the NPN and PNP versions, respectively.

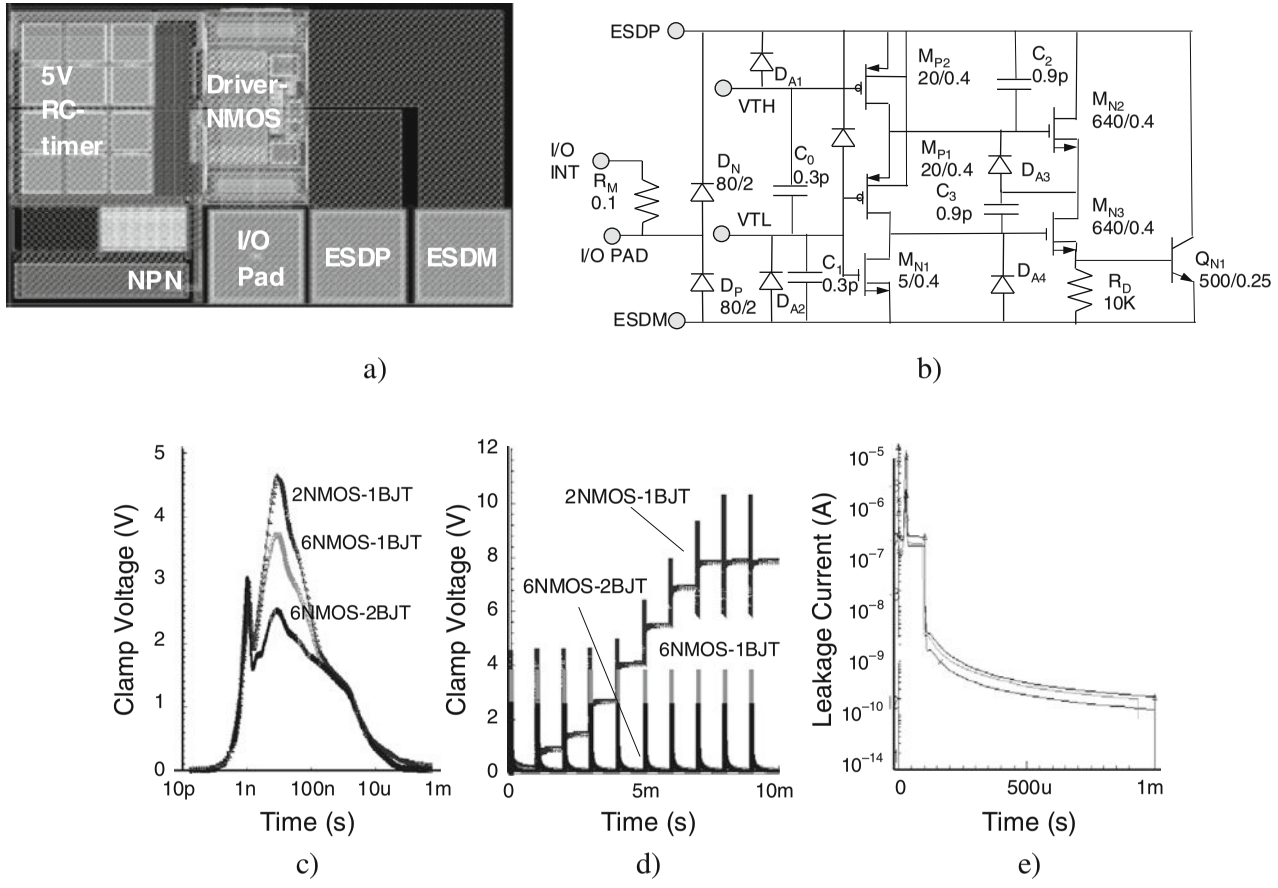


Fig. 5.12 Layout view (**a**) and circuit diagram (**b**) for the 5 V NPN-based slave clamp and the results of circuit simulation analysis for the single pulse waveform (**c**), multiple zap sequence (**d**), and power supply ramp (**e**) for different clamp component width scalings

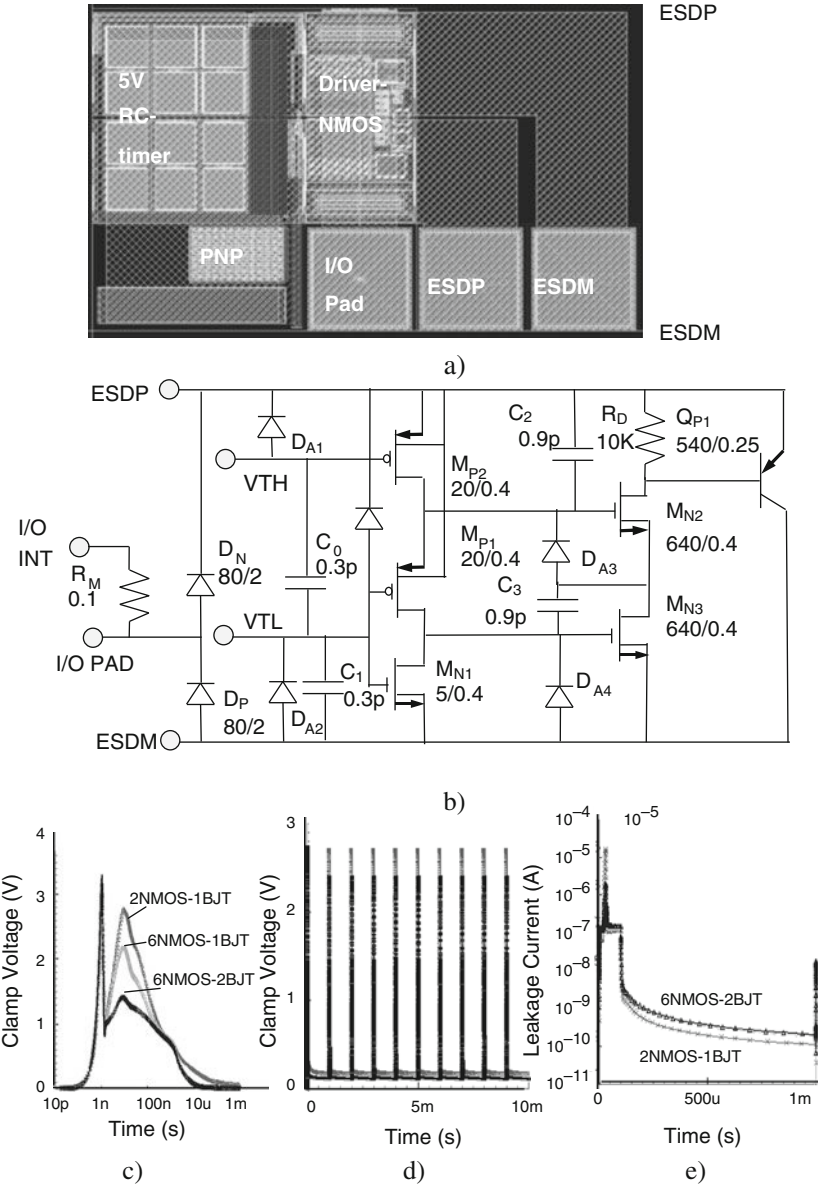


Fig. 5.13 Layout view (**a**) and circuit diagram (**b**) for the 5 V PNP-based slave clamp and the results of circuit simulation analysis for the single pulse waveform (**c**), multiple zap sequence (**d**), and power supply ramp (**e**) for different clamp component width scalings

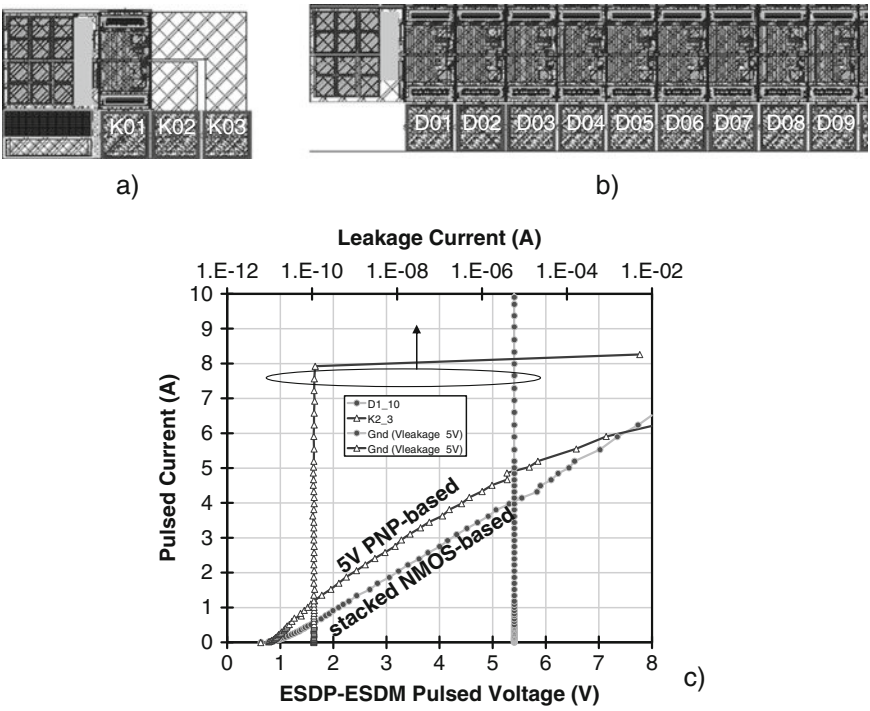


Fig. 5.14 Comparison of the layout footprint for 5 V ESD protection solutions: 1-clamp PNP BJT (**a**) and 10-clamp stacked NMOS (**b**). Comparison of the experimental TLP characteristics for ESDP–ESDM pad combination for the cluster of 10 stacked 3 V NMOS clamps and PNP-based active clamp (**c**)

5.1.5 Bipolar Differential Input Protection

Rail-based ESD protection is one of the most typical solutions used for power amplifiers. In the case of bipolar differential input, the typical emitter–base breakdown voltage with an open collector (BVEBO) is ~3–6 V. Thus, the base–emitter junction is usually rather sensitive to voltage overstress and requires appropriate voltage limiting to avoid damage.

Since the absolute maximum voltage at the input pins is relatively low, diode clamps are used to protect the differential BJT input pins. The voltage limiting across the base–emitter junction can usually be implemented in two ways, using either large differential diode clamp or the two-stage network approach with a small differential diode clamp. The final choice depends on the circuit specifications.

If a larger amplitude of differential voltage is required (Fig. 5.16), up to three diodes can be stacked in the clamp. When such a diode clamp is used as a primary ESD path between the inputs during an ESD event, a 1.5 A current through each diode results in a voltage drop. The diodes should be appropriately oversized to

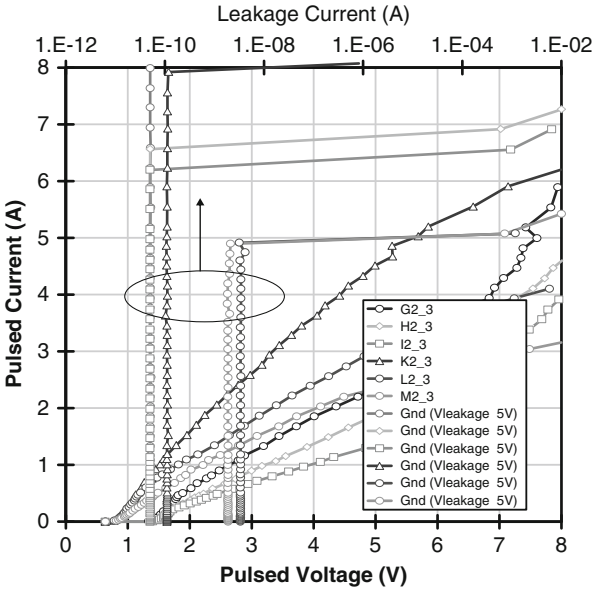


Fig. 5.15 Comparison of the experimental TLP characteristics for the ESDP–ESDM pin combination for the different versions for NPN and PNP BJT-based clamps. (G) 1080 × 0.24 μm^2 NPN BJT-based clamp; (H) 540 × 0.24 μm 2 NPN BJT; (I) clamp "H" with a three times reduction of the output-stacked NMOS device width in the driver; (K) 1080 × 0.24 μm^2 PNP clamp; (L) 540 × 0.24 μm^2 PNP clamp; (M) clamps "L" with a three times reduction of the output-stacked NMOS device width in the driver

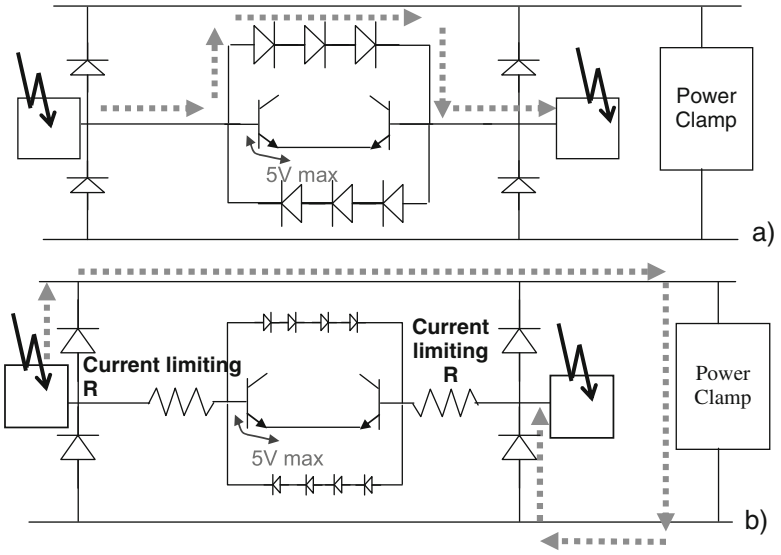


Fig. 5.16 Bipolar differential input protection without a current-limiting resistor at the pads, a large ESD diode clamp (**a**); with current-limiting resistors at the pads, a small differential diode clamp (**b**)

provide a high-current voltage drop of less than 1.5 V, which will keep the total clamp voltage drop within 4.5 V. This approach is mainly used when the inputs cannot tolerate additional resistors due to the degradation of circuit performance.

A more sofisticated approach can be implemented if a current-limiting resistor can be included in the circuit (Fig. 5.16b). In this case, much smaller diodes can be used in the back-to-back stacked diode clamp to provide a similar clamping voltage at the inputs. The resistor value must be sufficient to limit the current when the voltage between the pads is above the power supply level. A typical resistor value is 150 Ω and can be precisely calculated using BJT compact models and circuit simulation.

The number of diodes in the differential diode clamp depends on the desired input signal amplitude. However, even if the signal amplitude is small, in most practical cases it is desirable to have at least two stacked diodes with shallow diffusion connected to the pins to reduce the parasitic effect of the substrate junction and balance the leakage current. For example, stacked diodes can be constructed by stacking the p^+ -Nwell and n^+-Pwell diodes, connecting the well tap diffusions in the floating node.

5.1.6 Bipolar Output Protection

In the rail-based ESD network, bipolar output driver protection can be implemented using a standard NPN BJT clamp and one of the voltage reference techniques in the collector–base circuit (Chapter 4). This measure is effective if the output BJT devices remain under low base current conditions during ESD stress and thereby provide a high critical voltage. If the critical voltage is above one that corresponds to an open base voltage, then an open base reference can be used in the power clamp (BVCEO) (Fig. 5.17).

However, if the base coupling is high, then the NPN collector–emitter current path should be protected in a different way. This statement is true unless the NPN emitter area is sufficient in providing self-protection of the output devices. Usually, self-protection is possible at an NPN emitter size of ~500 μm^2.

The worst-case scenario is realized when there is an additional capacitor connected between the base and collector circuit nodes. When the output pin connected

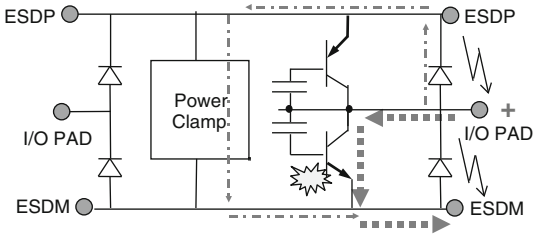


Fig. 5.17 Two ESD paths through the circuit illustrating the possible destruction of the output driver if the ESD current through the device is not limited

to the collector is stressed, the capacitor provides a positive base current with a corresponding reduction of the triggering voltage, at a small multiplication factor (Chapter 3). As a result, all of the ESD current is directed through the NPN (Fig. 5.17).

A countermeasure to suppress this effect is to add another current-limiting resistor. To protect the output pad in this case, it is necessary to ballast the collector with a sufficient resistance or to ballast the emitter with a sufficient polyresistance to limit the current below a safe level.

The choice depends on the size of the bipolar device. These options can both be implemented by having only a single resistor for each bipolar device or by implementing separate local resistors at each segment (finger).

Due to a high current provided by the BJT devices, in most practical cases, the total ballasting resistance should be within $\sim 1\ \Omega$. This value often insignificantly impacts output performance, especially in the case of high-current output. The most practical design in the case of power amplifiers takes into account heat generation balance and mutual heating. In this case, the output drivers NPN and PNP are drawn as multiple transistors, with polyresistors ballasting at each individual emitter. The polyresistance per emitter value could be rather high, for example, $100\ \Omega$ per device finger.

In an isolated process, another efficient way to protect the output stage is by adding a Pwell resistor between the pad and driver output. In this case, current limiting is provided for both the NPN and PNP output drivers. The Pwell resistor is preferred over the polyresistor, as it might benefit from some current saturation under ESD stress – resulting in a higher effective resistance.

5.1.7 CMOS Input and Output Protection

In the case of BiCMOS circuits, both bipolar and CMOS outputs and inputs should be protected. For the CMOS domain, rail-based ESD protection can be realized using automatically voltage-tolerant CMOS-based power clamps. This can be done by implementation of either the active CMOS clamp or one of the CMOS-based snapback clamps, for example, the snapback NMOS clamp.

The snapback NMOS clamp is usually preferable, since FOX devices produce a too high triggering voltage to be suitable for rail-based protection. At the same time, SCR devices generally produce a relatively low holding voltage below the power supply level, thus creating potential conditions for the transient latch-up scenario.

Therefore, snapback NMOS protection is usually preferrable as the self-aligned option that is the least sensitive to process variation. However, since the NMOS structures in the snapback and output devices are rather similar, there are several important considerations that should be taken into account.

The first problem is in the grounded-gate power clamp, which provides relatively high triggering and holding voltages. While the triggering voltage can be reduced by the dynamic coupling of the gate, the high holding voltage should be accounted for in the total voltage drop.

Based upon the SOA of NMOS and PMOS, the most critical factor is the snapback of the NMOS output driver, while the PMOS device can withstand a much higher voltage level before burnout.

If rail-based protection is realized with the snapback NMOS power clamp for the PAD–GND ESD zap, the current may be directed through the output NMOS rather than power clamp, due to the additional voltage drop on the diode and metallization.

The first step in improvement is to add a local snapback NMOS clamp between the PAD and GND. This solution will at least eliminate the additional voltage drop on the upper diode and metallization (Fig. 5.18a). However, even during ESD, the transient triggering voltage of this additional voltage clamp will compete with the triggering voltage of the output NMOS, which might have rather large gate coupling.

If the open drain protection is required, then a resistor at the PAD can be used to limit the current into the output NMOS device. The value of the resistor can be calculated using the measured value of the critical avalanche current that is supported by the NMOS device reversibly prior snapback.

In case of low gate potential of the output NMOS, the lowest avalanche breakdown current density per micron gate width that can be expected is $j_{CR} \sim 0.02$–0.05 mA/μm. However, in this case, the triggering voltage of the NMOS is high and greater than that of the local clamp. If gate coupling in the NMOS driver is high, then the device can be expected to operate above the threshold. In this case, the triggering voltage will correspondingly decrease down to the range of the holding voltage.

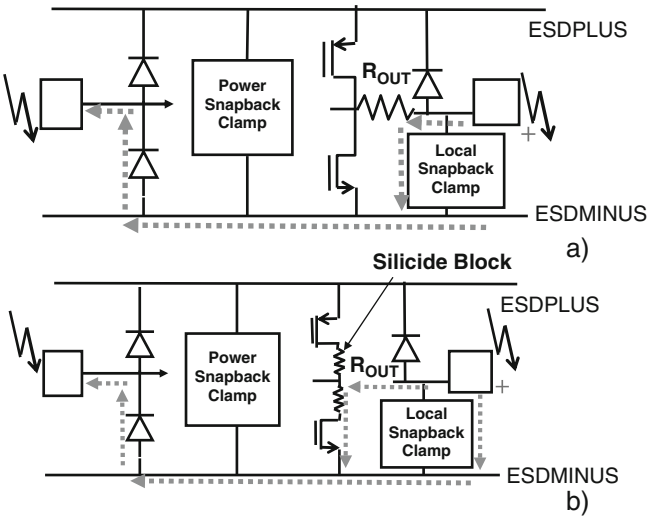


Fig. 5.18 Simplified circuit diagram of an NMOS output pad with a current-limiting resistor and an additional local clamp for ESD protection (**a**); self-protecting output based upon NMOS with a drain ballasting region (**b**)

An open channel current j_{CR} can be used to roughly estimate the required decoupling resistor between the output node and pad. The resistor should provide a voltage drop above the difference of the triggering voltage V_{T1} for the local snapback NMOS clamp and the holding voltage V_H of the snapback NMOS. The last is considered a worst-case scenario. Thus, for an NMOS driver with a total gate width W_G, the resistor value should be above the estimated value:

$$R_{OUT} \sim (V_{T1} - V_H) / (W_G \times j_{CR}) . \tag{4.1}$$

In a 5 V CMOS process with $V_{T1} = 10$ V, $V_H = 6$ V, $W_G = 500$ μm, and $j_{CR} = 0.1$ mA/μm, a safe value for R_{OUT} is >80 Ω. In the case of practical design, one of the possible sources of mistake of such an estimate for relatively large NMOS arrays is a non-uniform transient current distribution that may result in local triggering of the array below the estimated total current level. The total current level can also be impacted by local process-related inhomogeneities formed in the array.

An alternative solution is illustrated in Fig. 5.18b, where the output driver NMOS is modified according to the same rules that are used to design a snapback NMOS cell: a ballasting region is added into the device to produce reversible snapback operation. For the same purpose, the design of the device also includes maximum well tap to source spacing.

As a result of such changes, the output NMOS driver can reversibly operate in snapback mode with drain current saturation. In this case, the design can support two variations, depending on the size of the output driver. If the device width exceeds ~200 μm, then if properly designed, it can conduct an ESD current level and thus a local clamp at the pad is not required. In the opposite case, the clamp can conduct the excess ESD current. The only rule in implementing such an operation is to provide a sufficient saturation region of the snapback characteristics to make sure that the triggering voltage level of the local clamp is achieved before snapback and burnout of the output NMOS.

The self-protecting NMOS solution is very efficient. In most cases, the silicide blocking region or drain-contact-to-gate region for non-silicided processes of only a 2–6 μm long is usually required. Thus, the total accumulated resistance in the drain path is rather low. For example, in the case of sheet resistance of ~100 Ω/square for an n^+ -diffusion region, even a small NMOS device with a 20 μm width accumulates additional drain resistance of only 10 Ω for 2 μm long drain ballasting region. At the same time, such a saturation region is capable of providing a sufficient saturation to engage the local ESD clamp before critical conditions for burnout.

5.1.8 Array-Level Consideration

Certain changes can be made to the NMOS output device layout in order to support the chosen strategy of ESD protection.

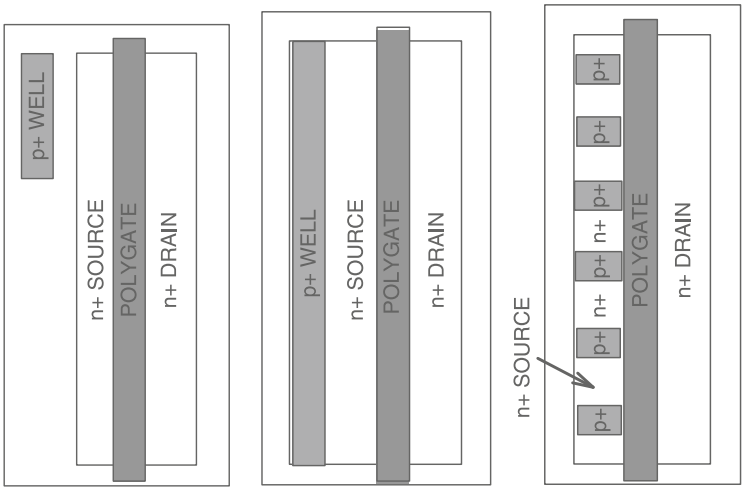


Fig. 5.19 Versions of output NMOS array design with an arbitrarily placed non-butted Pwell contact diffusion region (**a**), with butted diffusion region (**b**), and with interdigitated PLUS region (**c**)

If the output NMOS is expected to work in snapback, then, in addition to the drain ballasting region, an appropriate maximum spacing between the Pwell contact and the n^+ -source diffusion should be provided in order to improve the snapback operation mode with the higher internal resistance of the parasitic NPN. Device topology for this case is presented in Fig. 5.19a.

However, if the protection relies on the local snapback clamp, then increase of the triggering current of the NMOS driver before snapback is desirable. In some cases, a greater gate length can be selected, which will also provide a higher triggering voltage.

The purpose of modification is to provide the lowest possible internal substrate resistance to reduce the base resistance of the parasitic NPN. The simplest solution is to provide each array finger with individual P-WELL contacts butted to the source (Fig. 5.19b).

One of the topologies that provide the lowest substrate resistance with small footprint penalties for an NMOS array can be implemented by the introduction of interdigitated p^+ -regions, as shown in Fig. 5.19c. In this case, contacts to the p^+ -regions are shared with the source contacts, and the internal base resistance of the parasitic NPN is minimal.

A similar measure also improves the parameters of NMOS arrays used in output circuits that work on the reactive load in switching voltage regulators. In this case, during the load mismatch short-term voltage overstress of the drain, the high voltage may result in an accidental triggering and burnout of the output NMOS device.

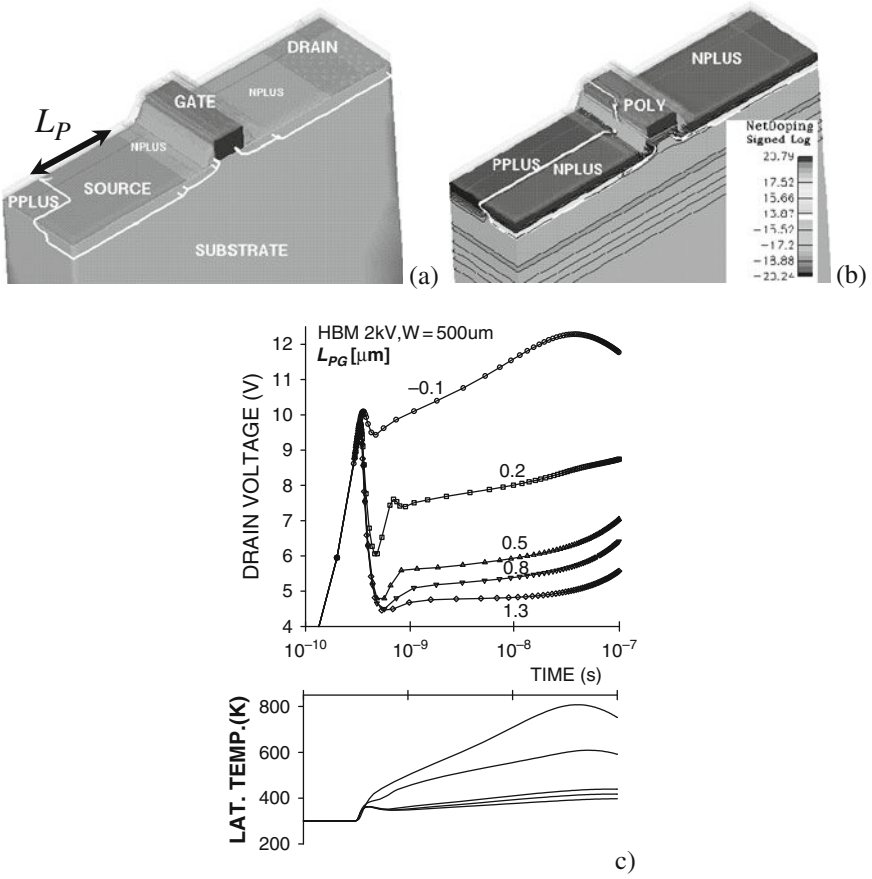


Fig. 5.20 Three-dimensional segment of the power array that represents the physical effect of the p^+ -contact as a function of p^+ -to-polyspacing (L_P) (**a**), the example of net doping profiles for $L_P = –0.1$ μm (**b**), and the results of transient mixed-mode 3D analysis for ESD pulse operation of this segment, with area scaling factor 500 at 2 kV HBM. The dependance of waveforms of the drain voltage upon L_P (**c**) demonstrate the high holding voltage of the array

Summary of a 3D numerical simulation of this physical effect is presented in Fig. 5.20. As a physical representation of a periodic array, a 1 μm wide 3D segment of an NMOS device was selected (Fig. 5.20a). The segment was numerically simulated with mirror boundary conditions. Variable spacing between the embedded P-region and polygate L_P has been used to demonstrate the effect. An extreme case is obtained when the p^+ implant mask overlaps with the polygate region $L_P = –0.1$ μm in the device (Fig. 5.20b).

The calculated transient waveforms for a 2 kV HBM operation of the ESD HBM in mixed mode are presented in Fig. 5.20c for different values of the L_P parameter.

At relatively large p^+ -to-polygate spacing, the device exhibits typical snapback operation with the holding voltage equal to ~4.5 V. With L_P reduction, the holding voltage of the device increases and, for the case of p^+ -implant to polyoverlap, the device provides a practically non-snapback operation with a bell-shaped waveform. This effect is the result of high efficiency in the collection of the holes from the source–drain spacing, which produces a corresponding misbalance between the electron and hole currents.

5.1.9 Concept of Two-Stage Protection

Practically, the two-stage approach has already been illustrated above each time separation resistors are introduced between the protected pin and active circuitry. This section just provides a formal summary of the two-stage approach.

5.1.9.1 CMOS Input

Two-stage ESD protection is a main approach for the protection of CMOS inputs. The CMOS input is very sensitive to the ESD discharge, due to the relatively low voltage and current of gate dielectric breakdown. The gate dielectric breakdown is the major limiting factor for the passing level of charged device model (CDM) testing for high pin count products implemented in scaled CMOS processes with gate lengths 0.18 μm and below. At the same time, in the case of 5 V processes, no special consideration is required because the pulse breakdown voltage of the gate oxide is above the triggering voltage of the snapback clamps. The rail-based solution is exceptionally robust in the case of an active power clamp.

Additional input resistors can be used to separate the inputs and realize two-stage ESD protection using a set of diodes. The principle of the two-stage network is widely used in system-level ESD protection (Chapter 8). In this case, the first stage is formed by the discrete suppressor component.

In principle, the two-stage network consists of a combination of two ESD clamps separated by a resistive current path or delay component. The resistor-based two-stage network is presented in Fig. 5.21. In this case, the first clamp limits the operational voltage at V_P, while the much smaller second-stage clamp limits the voltage in the internal node at V_I. The second-stage resistor R_S is selected depending on the second-stage clamp current I_{C2}: $R_S = (V_P – V_I)/I_{C2}$.

The second-stage clamp is very efficient in filtering short pulse overstress typical for CDM events due to the $R_S C_I$ time constant; C_I is the equivalent capacitance of the internal node.

The first clamp directly interfaces with the input PAD. It limits the voltage and conducts the majority of ESD current, but is not designed to precisely target the ESD protection window.

The advantage of the first-stage clamp is that usually the clamp can reversibly conduct above approximately three times larger current at an additional voltage drop above the desirable voltage limit within ESD protection window. For example, n^+ -Pwell and p^+ -Nwell diodes implemented within the minimum process design rules

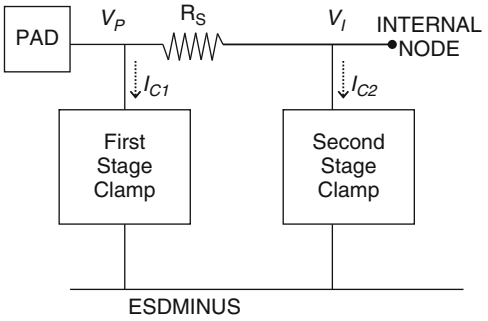


Fig. 5.21 Two-stage ESD protection for CMOS input

of a 0.5 μm process can deliver several times more current at an ~5 V drop than at a 1 V voltage drop.

Often, implementation of a much bigger diode cell is a rather poor alternative due to its accompanied increase in parasitic capacitance, leakage, and noise.

Thus, the first-stage clamp is used to transform the ESD pulse into a waveform with a much lower voltage amplitude of ~5–7 V. Then, much smaller second-stage diodes limit the voltage to ~1 V at the current level limited by the stage resistor R_S, which provides the remaining 4 V voltage drop.

Often, it is advantageous to implement R_S as a saturation resistor. However, due to the resulting resistor parasitics in high-speed I/O, a polyresistor provides a more predictable result.

5.1.9.2 Diode-Based Compact Two-Stage ESD Protection Circuit

Principles similar to the component-level ESD protection network can be implemented at the device level. A compact solution is presented as a three-terminal diode structure (Fig. 5.22). It incorporates two-sided diodes that form a three-terminal structure. One side of the diode is used for power ESD operation (a power diode connected to the I/O pad) and the other side just holds the lower potential in the internal circuitry (a potential diode connected to the internal circuitry). Both internal diodes share the same well, thus creating a two-stage ESD protection circuit with a corresponding "built-in" internal well resistor. A key important feature of the diode structure is the internal potential sharing during ESD operation.

Due to the realized internal potential sharing, a potential limiting is possible on the internal circuitry. The ESD protection of the I/O pin can be realized using a corresponding circuit presented in Fig. 5.23.

This approach is advantageous for a small footprint, integrated high-speed I/O compatibility due to the reduced equivalent capacitance.

Operation of the two-stage components can be analyzed using numerical device simulation of the cross sections generated by physical process simulation tool (Fig. 5.24a, b). During the high-current ESD event and the high-voltage drop on the pad side contact, the internal node contact potential remains rather low – 1–2 V only.

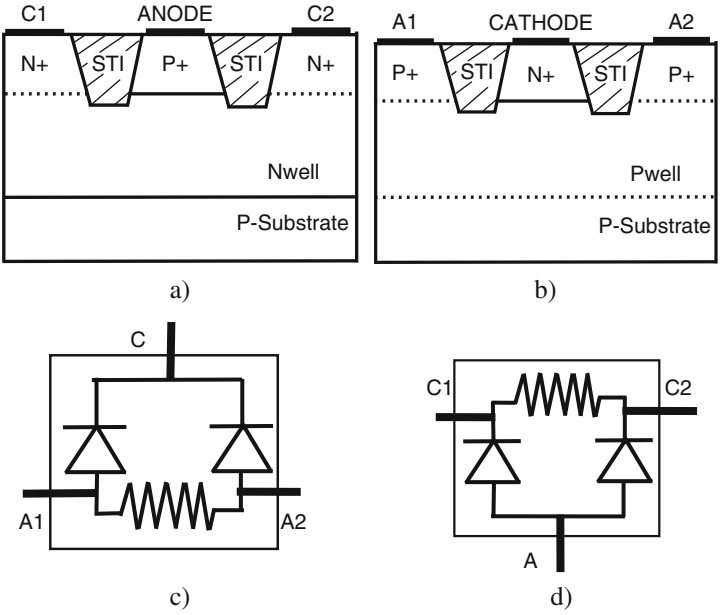


Fig. 5.22 Three-terminal diode structures for N-well (**a**) and P-well (**b**) replacement CMOS diodes used in the suggested ESD protection circuit for I/O cells. The simplified equivalent circuits of the three-terminal diode structures (**c**, **d**)

The next level of numerical validation can be achieved by the thermal-coupled mixed-mode simulation of different ESD HBM pulse levels (Fig. 5.24e, f). According to these results, ESD operation of the diode structures demonstrates that a maximum lattice temperature near critical values for the structure melting is reached under the power diode voltage level of ~6 V. At the same time, the internal terminal holds the voltage level below 2 V in this range (Fig. 5.24e, f).

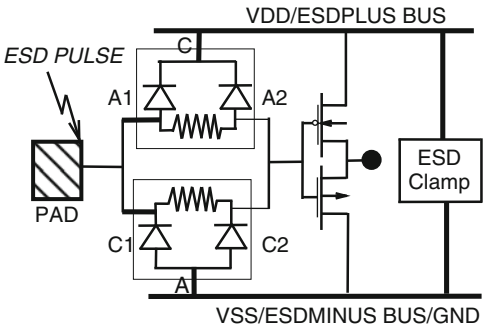


Fig. 5.23 I/O circuit ESD protection concept using the three-terminal diode structures for rail-based protection

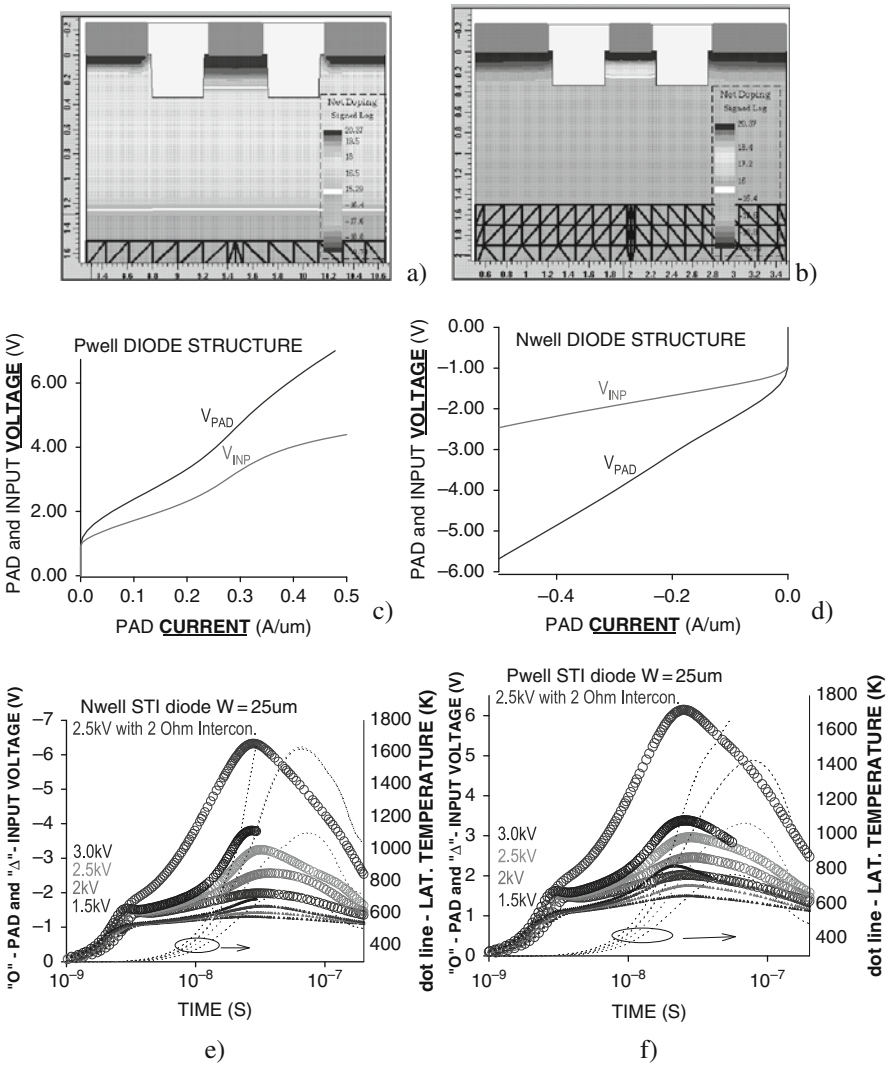


Fig. 5.24 Cross section of the N-well (**a**) and P-well (**b**) diode structures obtained from process simulation and corresponding *I–V* dependencies with the "INPUT" terminal potential (**c**, **d**) and results of numerical experiments for different amplitude HBM ESD operations: The pad and input voltages and maximum temperature for the conventional and proposed triggering clamps on the basis N-well (**e**) and P-well (**f**) two-stage diodes, respectively

5.1.9.3 Two-Stage ESD Protection Circuit for BJT Base

A similar protection approach can be implemented for the sensitive-to-voltage overstress BJT input. The cross section for the BJT device and input circuit is presented in Fig. 5.25a, d. This solution is an alternative to the diode cell protection (Fig. 5.25b).

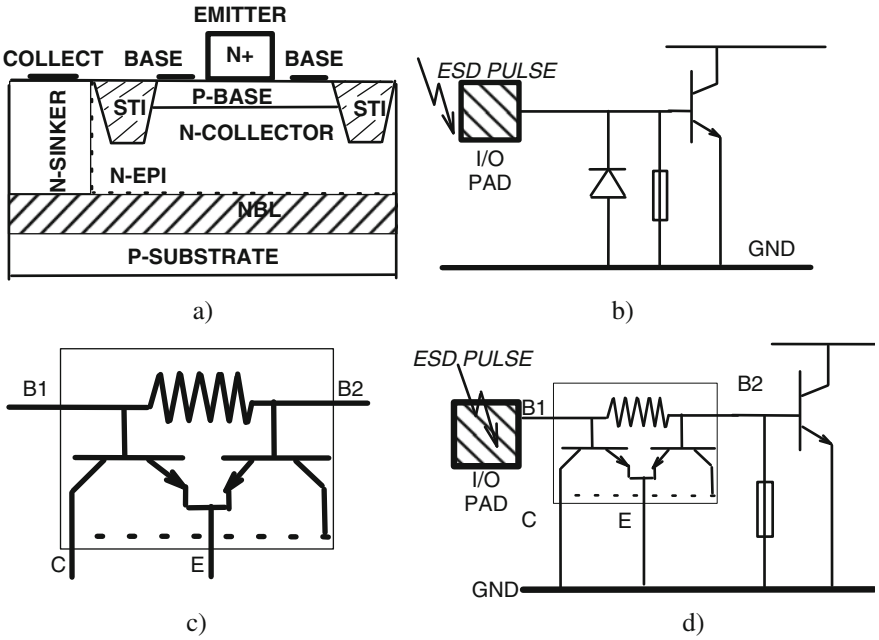


Fig. 5.25 Two-stage protection of the base–emitter junction: input NPN BJT cross section (**a**); typical protection circuit with diode (**b**); equivalent circuit for double base configuration (**c**); and proposed operation circuit (**d**)

To realize two-stage protection, two base contact regions from the opposite side of the emitter can be used to implement two-stage component (Fig. 5.25c) and the corresponding network (Fig. 5.25d). In this case, the base sheet resistance under the emitter region plays the role of the second-stage resistor clamp component. This internal base resistance is relatively small.

However, according to numerical simulation results, the effect is substantial. The numerical simulation has been performed using process-generated cross sections for a SiGe NPN device (Fig. 5.26a). According to the numerical analysis, the input side of the base remains at a potential of ~1 V in the wide range of the input ESD current even up to temperatures above the critical level for local device melting.

5.1.9.4 Two-Stage Network with Snapback NMOS

Similar principles can be applied to the snapback NMOS protection structure. The solution requires an additional potential contact that will sense the voltage from some location in the ballasting region.

The ballasting region of the NMOS already represents a diffusion resistor that operates in deep saturation mode in the case of snapback. Thus, if the internal node potential is sensed closer to the gate of the NMOS device, the voltage limiting in the internal node might be achieved at a significantly lower level.

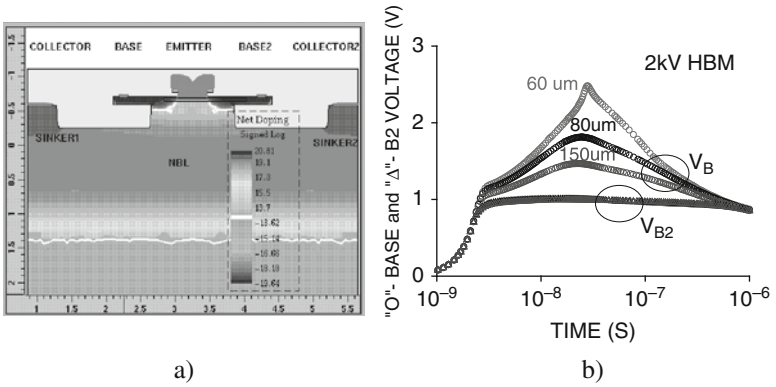


Fig. 5.26 Process simulation cross section of SiGe NPN BJT (**a**) and mixed-mode simulation results for voltage at pad side V_B and internal node side base contacts V_{B2} for 2 kV HBM ESD pulse for different structure width (**b**)

Thus, the separation of the "input" and "output" voltages is realized on the device level in the snapback NMOS structure. In order to make this possible, an extra "potential-sensing" terminal is introduced between the drain metal (silicide) and the gate (shown in Fig. 5.27a, b). Figure 5.27c presents the suggested local clamp for ESD protection.

During ESD operation with a high-current pulse, the voltage drop on the ballasting region results in a corresponding electric potential drop in the drain ballasting

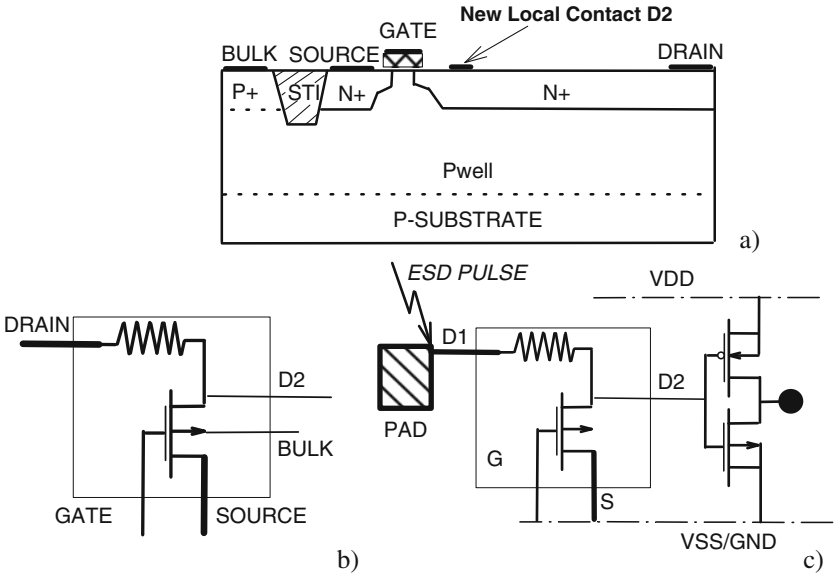


Fig. 5.27 Suggested GGNMOS with local potential terminal (D2) (**a**), its simplified equivalent circuit (**b**), and suggested local clamp implementation (**c**)

region. Thus, the voltage at the added contact depends on its position inside ballasting region and, in general, is significantly lower than that at the drain contact and, at least, does not include the voltage drop on the interconnects and contacts. A key important feature of the structure is the internal potential distribution during power (ESD) operation and is illustrated below using TCAD tools.

A side effect (or trade-off) of the implementation of this circuit is the resistor between the pad and the internal circuitry. However, due to a rather low n^+ -composite sheet resistance of <80 Ω/square, the estimated resistor value is usually negligible below 1–10 Ω, depending on device dimensions.

At the same time, the "benefits" from the presented solution are rather significant. Based upon the distribution of the electric potential inside the snapback NMOS structure during high-current ESD pulse, low voltages can be sensed (Fig. 5.27) closer to the gate region. Ultimately, the voltage level is comparable with the holding voltage. If the internal circuitry is connected to such terminal, it remains protected at a low 5–6 V level independently from the voltage drop on the clamp at high current, thus eliminating the risk of gate oxide damage (Fig. 5.28).

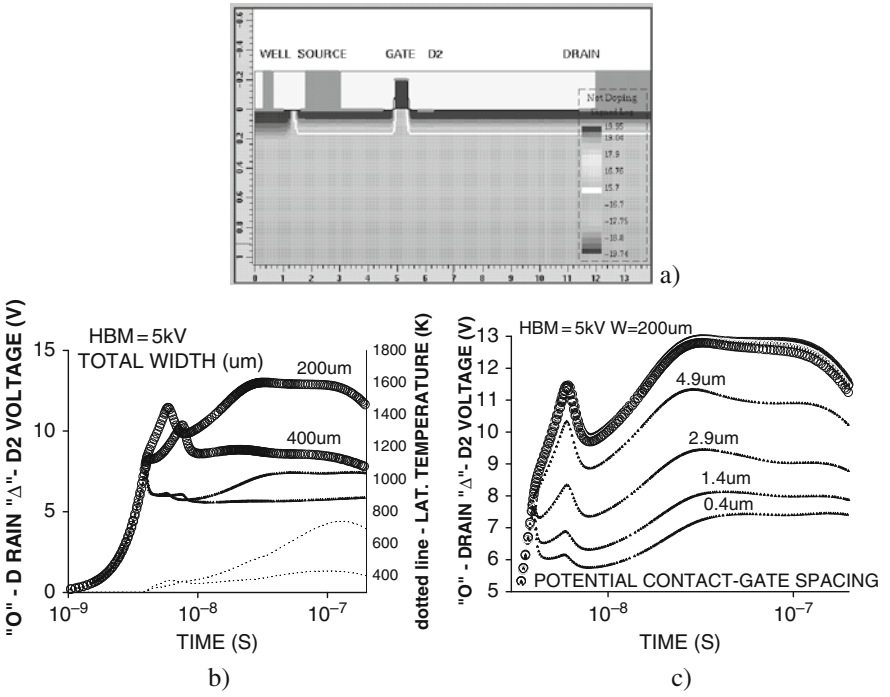


Fig. 5.28 Process simulation cross section of the snapback NMOS (**a**); results of the numerical simulation for a 5 kV HBM ESD pulse operation with different structure widths (**b**); and the voltage on internal circuit side contact "D2" depending on its position relative to the gate edge (**c**)

In practical design, in order to avoid interference of the sense contact metal with the ballasting region, a separate local n^+ -diffusion extension can be made at the side of each finger to accommodate an additional electrode spaced away from the active region.

5.2 Local Clamp-Based ESD Protection Network

5.2.1 Local ESD Protection

There are many analog circuit blocks that, in principle, cannot tolerate the rail-based ESD protection.

One of the typical examples of when a rail-based ESD protection network with diode clamps is not possible is the case of a back-drive compatible pin. The pin must be independent from power and thus the diode-to-ESDPLUS rail configuration is not possible.

A similar case is that of over-voltage input pins with thin film resistors. An example of an automotive application amplifier realized in a 5 V process with a 65 V-tolerant differential input is presented in [42].

Another example includes circuit pins with high dual-direction voltage tolerance. Protection solutions for such pins may not tolerate the lower diode-to-ground and may require excluding the upper diode-to-power supply rail as well.

In such cases, local ESD protection is used as either the only protection approach or combined with partial rail-based protection.

An example of a local ESD protection network is presented in Fig. 5.29. In principle, every pin in the network is protected by a local ESD clamp that provides a voltage tolerance suitable for the specified pin parameters.

The previously discussed two-stage ESD protection is similarly applicable for local clamp protection. One of the most typical ways to implement two-stage ESD

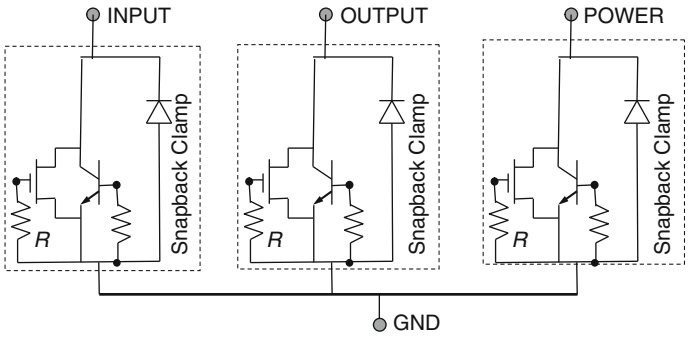


Fig. 5.29 Example of the four-pin local-based protection network

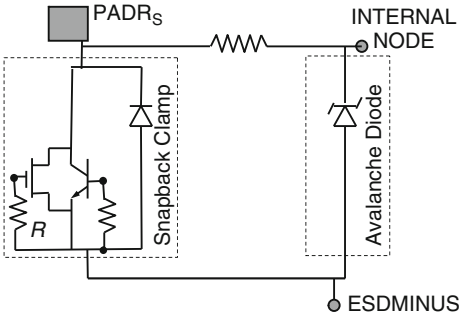


Fig. 5.30 Local-based network with the second-stage avalanche diode

protection is to include a snapback clamp separation resistor and a voltage-limiting avalanche diode (Fig. 5.30).

The complexity of local ESD protection is related to the generally unknown conditions of the control electrodes of the devices interfacing with the pad during ESD pulse. As a result of this the triggering voltage of the internal devices can become comparable to the triggering voltage characteristics of the ESD clamp.

5.2.2 Serial Data Line Pin Case Study

One of the cases for a CMOS pin protection can be illustrated with the serial data line (SDA) pin for the I^2 *C*-bus [105]. An open drain NMOS M_{DR} with a pull-up capacitor C_U between the gate and drain is required for circuit functionality (Fig. 5.31).

The open drain device M_{DR} is laid out with minimum spacing between the source and the well to increase the snapback triggering voltage. Nevertheless, the pull-up capacitor C_U ensures that the device remains in the on-state at the beginning of an ESD pulse before the ESD M_{ESD} triggers into high-conductivity mode. As a result, local burnout of the driver device M_{DR} is observed (Fig. 5.31).

A possible ESD protection solution for this pin is presented in Fig. 5.32. An additional resistor can be added between the two NMOS devices to increase the voltage level at the pad required for M_{DR} snapback (Fig. 5.32a). Additional measures can include engineering of the gate circuit in order to keep the M_{DR} gate at low potential during ESD stress until an ESD pulse sufficient to initiate snapback of the clamp, thus preventing the snapback of M_{DR}. Such a circuit can be realized by the addition of active components M_{K1} and M_{K2}, referenced to the power supply VDD node (Fig. 5.32b).

The "keep-off" circuit relies on the additional capacitance C_K during ESD stress, while VDD node is assumed at low potential. This results in an off-state M_{K1} and on-state M_{K2} during the transient rise time of the ESD pulse. After ESD NMOS clamp M_{ESD} is engaged, the low holding voltage guarantees non-snapback operation of M_{DR}.

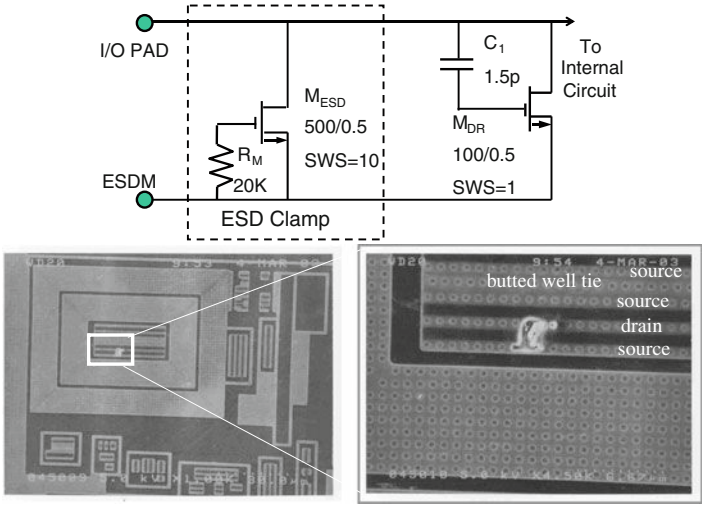


Fig. 5.31 Example of local protection for serial data line pin and the results of failure analysis

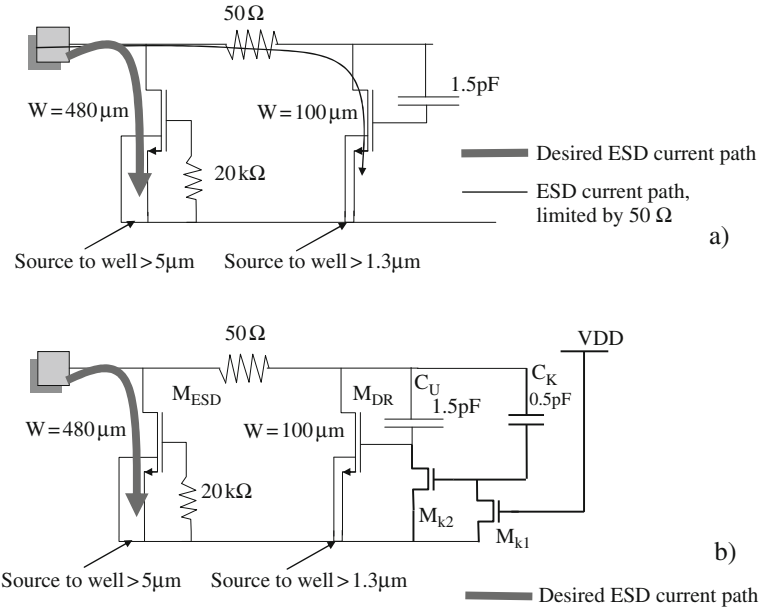


Fig. 5.32 Solution for SDA pin protection using a separation resistor (**a**) and additional "keep-off" circuit (**b**)

At normal operation with VDD signal, the keep-off circuit is disabled by VDD signal.

It is important that, unlike in active clamps, the required RC time constant provided by C_K and the parasitic resistance of M_{K1} is rather low (~10–20 ns). This

time is usually sufficient to trigger the snapback NMOS into avalanche–injection conductivity modulation mode.

5.2.3 Erase Pin Protection in EEPROM

One of the typical symptoms of inadequate protection of the "Erase" pin in EEPROM memory modules with an external high-voltage supply is functional failure. This failure is observed as a reprogramming of the memory cells by the ESD pulse, rather than ESD damage. This event has been experimentally observed, even though the programming normally happens only at some specific voltage combination on the "erase" and "enable" pins. However, during transient ESD events, some pins and internal nodes can be floating, thus creating conditions for functional failure due to reprogramming.

Thus, an ESD protection solution is required not only to provide an appropriate voltage limiting during the initial part of the ESD stress but also to make sure there is no high voltage left at the pin as a result of the stress.

For a 0.5 μm CMOS process, the typical erase pin voltage level is 12–20 V, depending on gate oxide tunneling characteristics.

In the case of a small footprint solution, first-stage protection includes a corresponding high-voltage snapback device. However, a single snapback device is insufficient, due to the relatively high residual voltage after snapback [62]. Therefore, a proper solution requires an additional high-voltage device. The first version of the local ESD protection includes snapback SCR device (Fig. 5.33a). The solution relies on the additional voltage clamping provided by the snapback NMOS, in addition to the SCR.

However, the success of the application of this solution depends on the size of the memory array. In a single cell, the residual voltage is still rather high (Fig. 5.34), and this leads to reprogramming of the single cell array. The most important experimental fact is that there is no discharge path in the pin after turn-off. The HBM waveform measurements demonstrate long-term storing of the residual voltage after ESD stress (Fig. 5.34).

The circuit has been modified to eliminate the high residual voltage on the erase pin after the ESD event (Fig. 5.33b).

In order to add the discharge current path required after SCR turn-off, an additional large resistor R_B is added between the pin and ground, and a large resistor R_2 is added between the pin and the memory. Finally, the second-stage clamp is redesigned to provide a current path after the SCR turns off (Fig. 5.33c).

A critical parameter is the equivalent capacitance of the connected memory cell. If only one cell is connected, the estimated capacitive load is ~20 fF. For the connected ~50 cells, the total capacitive load is ~1000 fF.

The capacitive load affects waveforms observed under ESD pulse conditions. In the case of the low capacitive load provided by a single cell, the residual voltage is above the erase level, while higher capacitive load of several cells provides a much lower residual voltage (Fig. 5.34). The following measures can be taken to guarantee

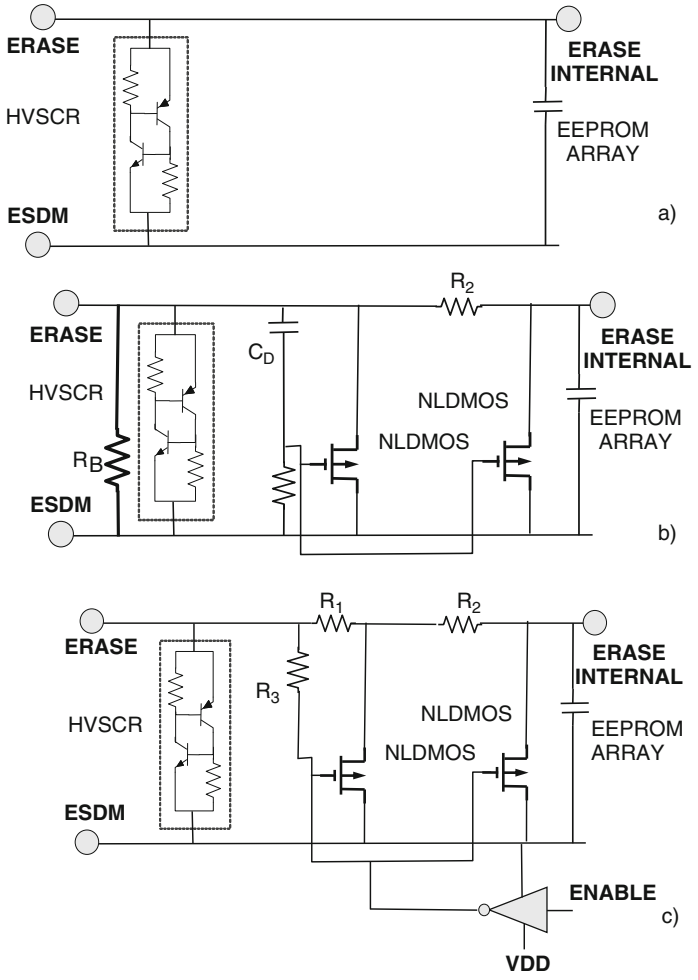


Fig. 5.33 Original (**a**) and modified (**b**, **c**) EEPROM erase pin ESD protection clamp circuit

low sensitivity to overstress: provide a discharge path and use clamps with a lower turn-on voltage.

An alternative, more compact circuit solution is presented in Fig. 5.35. This circuit is based on an NLDMOS-SCR with implemented gate circuit control. The principle of operation for this circuit can be explained as follows: the ESD device provides discharge of the majority of ESD current (I_{SN}) in snapback mode similar to a typical high-voltage NLDMOS SCR clamp with a grounded gate. However, due to the gate biasing circuit, the self-turn-off at a corresponding minimum critical current brings the device into a condition with channel current (I_{CH}) rather than into the low leakage state.

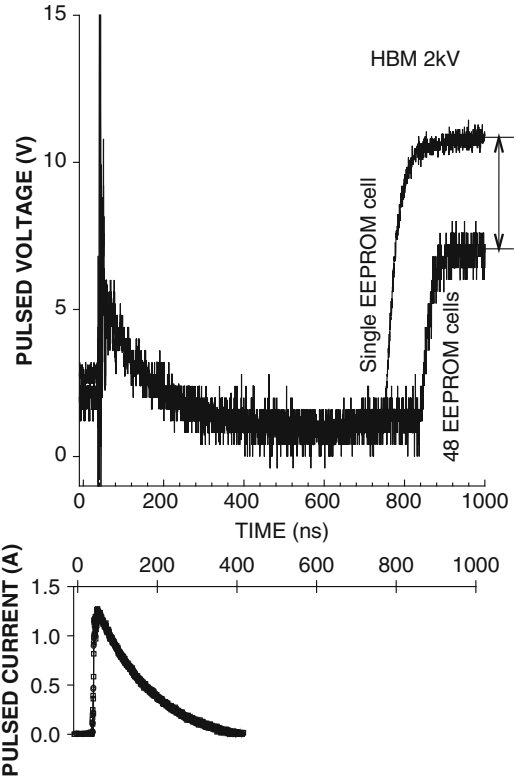


Fig. 5.34 Voltage and current waveform for the EEPROM with single cell and 48 cell arrays measured at 2 kV HBM

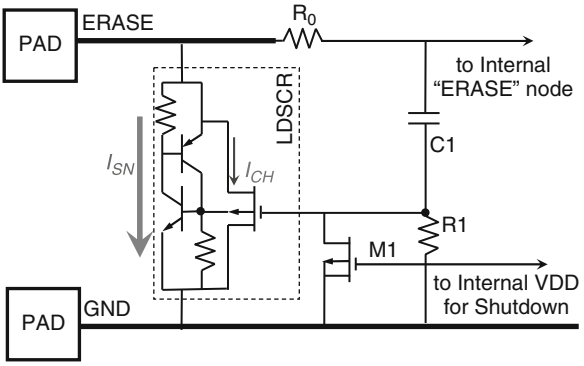


Fig. 5.35 Alternative small footprint EEPROM erase pin protection solution

The transient gate bias is supplied by the R_1C_1 circuit connected after the separation resistor R_0 ($\sim$100 K). The R_1C_1 time constant is less than $\sim$1 μs. During this additional time, the gate of the ESD device is held above the threshold V_T level. In these conditions, the channel current I_{CH} of the device provides the path

for the remaining current. The discharge continues until the internal node voltage drops below the threshold voltage V_T or until the RC delay time constant is reached. The critical minimum current in snapback mode can be as low as $I_{SN} \sim W \times 0.1$ mA/μm∼20 mA. After the majority of the ESD pulse is discharged through the open NLDMOS, the channel current is approximately constant for the rest of the pulse.

The circuit provides an appropriate functionality for a low-voltage Electrical Overstress (EOS) as well. In this case, at overstress below the snapback voltage level, the ESD device remains in the off-state while the LDMOS channel current path is supported.

At normal operation conditions, the VDD signal and small NMOS device M_1 keeps the ESD device gate at low potential. The capacitor C_1 is fully charged and no parasitic leakage current path is present.

Thus, this alternative clamp solves the EEPROM protection problem with a minimum set of components in a very space-efficient way. In this case, the ESD device is used both as a snapback and as an active component, providing operation in conductivity modulation and monopolar current conduction, respectively.

5.2.4 Local Protection of the Internal Pins

One of the most typical examples of the impact of ESD stress on the internal pins in high-voltage circuits is the internal voltage regulator of low-side control circuits, logic, and drivers for switching devices. Because of the relatively low impedance of internal regulators, a second-stage resistor cannot be used in order to implement two-stage protection. As a result, the regulators are usually rather sensitive to ESD damage and must be designed to withstand certain ESD current levels. Usually, the internally generated VDD voltage node requires an internal voltage clamp capable of handling high ESD current. Such a clamp is used even if the VDD node is not connected to the external pin at all.

The most typical locations of failure in internal voltage regulator design are the high-side high-power components. In the high-side PMOS-based regulator, namely the high-voltage PMOS fails (Fig. 5.36a). In this case, the gate clamping diode may not provide a substantial improvement (Fig. 5.36b).

A similar ESD current induced failure can be observed in bipolar and stacked regulators. In this case, due to the domino effect, damage in both stacked components might be observed (Fig. 5.36c).

ESD protection of the high-voltage PMOS used in VCC regulators (Fig. 5.36) is based upon the expected ESD path formation between the VIN and BOOST pins. In general, there are three current paths that could be realized for ESD pulse current (Fig. 5.37).

(i) The first current path is provided by the large switching NLDMOS controlled by the driver. The NLDMOS turns on for a short period of time due to the

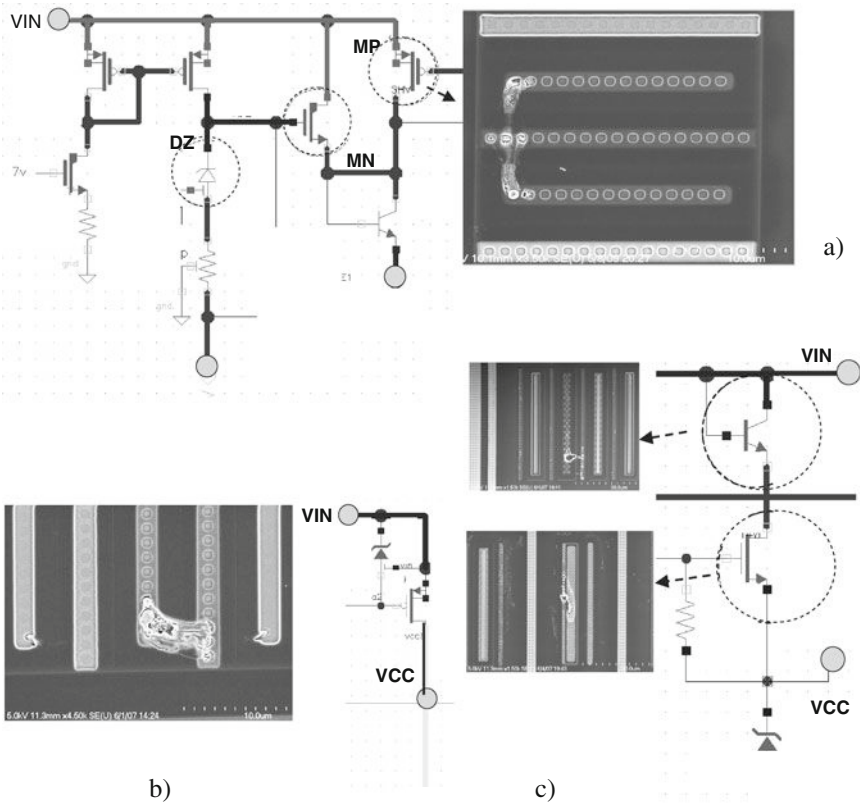


Fig. 5.36 Examples of the FA results for PMOS (**a**, **b**) and NPN BJT-based VCC regulators (**c**) [105]

drain–gate coupling that may be sufficient enough to discharge the ESD current. This current is directed to the switch pin SW, followed by the reverse current path through the BOOST-SW clamps. The clamp's reverse path ESD diode provides this remaining current path. This current path should not cause a failure in properly designed ESD protection clamps.

(ii) The second alternative current path is realized through the high-voltage ESD clamp to the power ground, followed by the ESD diode between the BOOST and ground. This current path is also reversible.

(iii) Finally, the third, and generally irreversible, current path is provided by the high-voltage PMOS itself and other stacked components of the regulator (for example, the NPN diode shown in Fig. 5.37) directly to BOOST. If this current path is dominant, the PMOS is exposed to failure. This current path explains the damage demonstrated by Fig. 5.36a, b.

Thus, in the case of a low level of array coupling and a small ESD protection window, ESD failures are related to the internal VCC regulator. The ESD protection

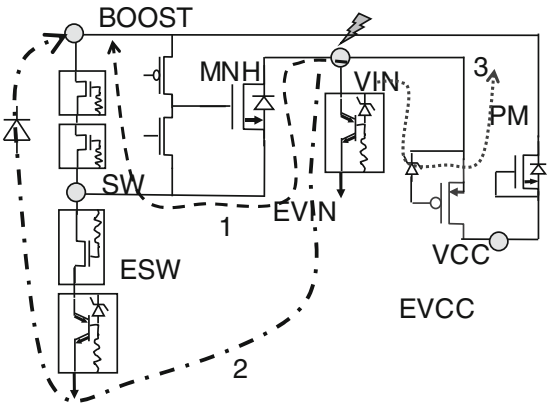


Fig. 5.37 Simplified circuit of the DC–DC voltage regulators with three possible ESD current path scenarios for VIN to BOOST ESD zap

network is very product-specific. There are several improvements that can be done to the circuit. However, in general, the circuit is still rather sensitive to process technology variation due to the following reasons: triggering voltage variation in the high-voltage ESD protection VIN clamp, dependence of the NLDMOS array self-protection capabilities upon layout, and internal circuit driver design. In addition, variation of the pulsed SOA of high-voltage PMOS and NLDMOS devices may provide an undesirable effect.

Design measures to improve the internal VCC regulator are based upon understanding of the current path alternatives listed above. One of the major assumptions is that in the general case, the current path can be shared and thus gradual improvement of all current path components may provide a desired effect.

The specific way to improve ESD protection depends on the dominant ESD current path realized in the particular product circuit. Perhaps, the first, most logical solution is to improve ESD protection by reducing the VIN clamping voltage to a lower level, thus limiting the voltage drop on the PMOS device. However, this adjustment of the parameters of the high-voltage ESD protection clamp might not be an option in case of relatively low SOA margins of the power-optimized components. In this case, the most robust approach is to implement ESD-oriented co-design of the VCC regulator, which provides a much higher pulsed absolute maximum voltage for the circuit pins. Design measures may include stacked or oversized components or application of less power-efficient components with higher voltage tolerance, if allowed by the process.

If the failed devices belong to the low side driver, then an additional internal power clamp, for example, the snapback NMOS clamp, may provide the solution. This clamp may replace the low-voltage avalanche diode between the VDD (VCC) and ground (Fig. 5.38). Even though the VCC node is internal, the additional EVCC clamp limits the critical voltage on the VCC node and provides a reverse current path.

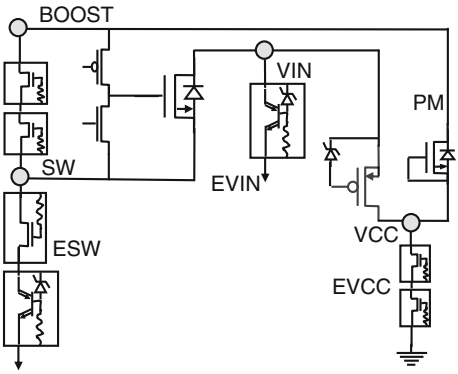


Fig. 5.38 Increase of local ESD protection robustness by application of the addition internal VCC to the power ground clamps

Another typical fix may not be related to the active area of the devices at all, but instead requires improvement of the metal connection inside the regulator to support the high current path. An example of the simple improvement effect of drain–source metallization is demonstrated in Fig. 5.39. In the layout of Fig. 5.39a, the device fails in the expected current crowding spot in the middle of the structure. A more uniform current distribution with metallization (Fig. 5.39b) eliminates the problem.

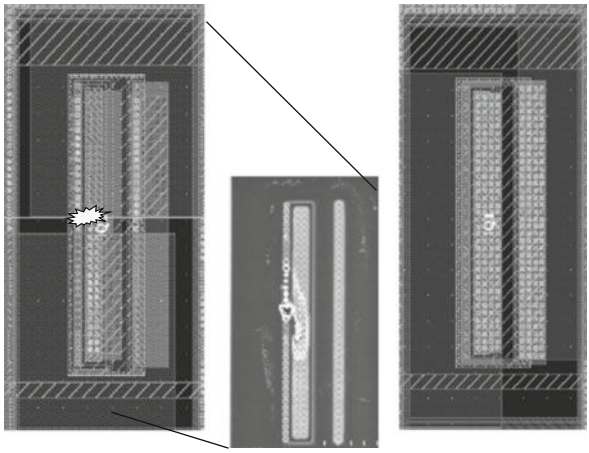


Fig. 5.39 Original metallization of the failed high-voltage PMOS with FA photo insert indicating the damage (**a**) and the successfully improved (**b**) device

5.2.5 Local Protection of the High-Speed I/O pins

High-speed I/O pins are usually co-designed with a printed circuit board (PCB) to work in a wide high-frequency band. In this case, the connection to the microchip

should be treated as a transmission line micro-strip network. In such a network, matching of the load cannot be neglected [106].

From the ESD solution point of view, signal propagation and reflections must be considered and matched to provide proper circuit operation with the implemented ESD protection network.

The relationship between the incident V_{inc} and reflected voltage signal V_{refl} in the transmission line, with impedance Z_0 and working on resistance R_l, is as follows:

$$V_{\text{refl}} = \left(\frac{R_l - Z_0}{R_l + Z_0}\right) V_{\text{inc}} = \Gamma V_{\text{inc}}, \tag{4.2}$$

where Γ is the reflection coefficient.

Three common cases of the transmission line on a matched load, open and short circuits are presented in Fig. 5.40.

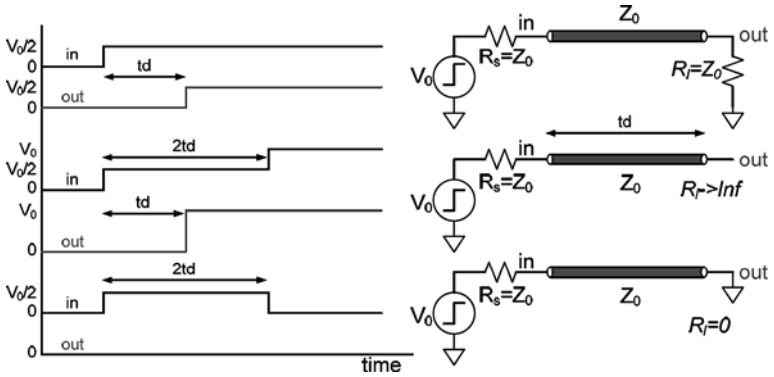


Fig. 5.40 Transmission line reflections [106]

ESD network design for high-speed voltage-mode drivers is generally compatible with NMOS or diode protection. Driver stacks can be difficult to configure with low-impedance drivers. Overall, driver resistance requirements limit the resistor value. Additional device width and parallel *R* are required to achive the appropriate passing level (Fig. 5.41).

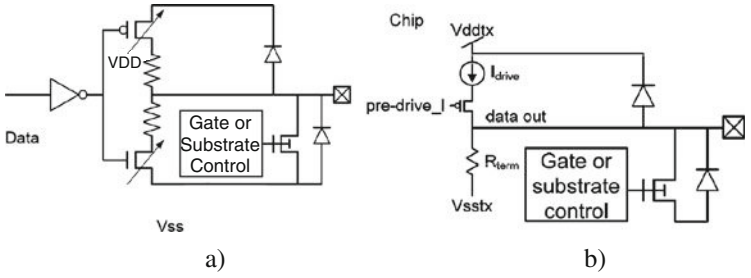


Fig. 5.41 Local ESD protection for voltage (**a**) and current (**b**) mode drivers with matching impedance resistors [106]

ESD for current mode drivers is also generally compatible with NMOS or diode protection. The current source in this case is usually a transistor stack, which provides higher ESD immunity. Termination R_{term} is usually calibrated, and transistor and resistor combinations must meet resistance limits.

Receiver design for a CMOS receiver with a standard inverter requires the balance of the NMOS and PMOS to ensure proper switching characteristics over the range of process variations, voltage, and temperature.

The major principle for successful high-speed circuit design is that the interconnects, driver, receiver, clocking, and ESD protection components must all function collectively. For greater speed, the interconnect channel and termination must be as clean as possible and must not provide any reflections, overshoots, or undershoots.

Receiver capacitance provides a significant limitation on signal integrity. Thus, the ESD protection solution requires a significant optimization of the ESD diodes and clamps to lower parasitics. This is achieved by metallization optimization, though even more sophisticated circuits involve auto-biasing of the ESD diodes to eliminate the parasitic capacitance effect. One of the methods is illustrated in Fig. 5.42.

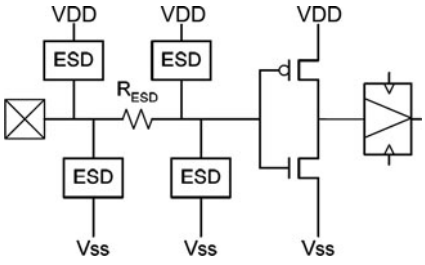


Fig. 5.42 ESD protection for a high-speed receiver with a two-stage matching resistor

5.3 ESD Network for Multiple Voltage Domains

5.3.1 Multiple Voltage Domains

When the analog circuit has different voltage domains, the ESD protection network can be constructed using several approaches that depend on circuit specifications.

The most typical example of a multiple voltage domain circuit is in the circuit with the digital, analog, and power domains. Since the ESD protection of a chip is required to support each pin-to-pin combination, all domains should be connected into a network.

Consider an example of analog and power domains. The inter-domain connection in the network is usually realized between the power and analog grounds. The simple reason for this arrangement is that the power supply voltages cannot

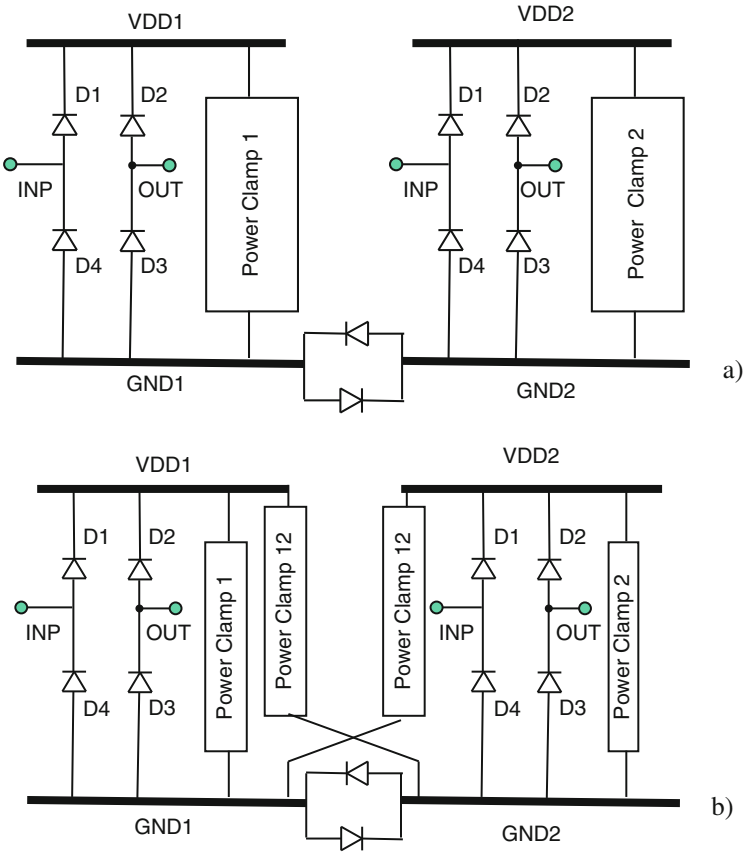


Fig. 5.43 ESD protection network for multiple voltage domains with a back-to-back diode clamp (**a**) and a cross-coupled power clamp (**b**)

be shared. To avoid cross-talk due to the potential drop on the ground busses, the grounds can be connected using a back-to-back diode clamp. This clamp decouples the ground buses if the voltage drop is less than the voltage drop on the forward-biased diodes and, at the same time, will provide a low voltage drop at ESD current level (Fig. 5.43a). In this case, the stress on the pins from the different domains will include the additional voltage drop on the diode.

An alternative solution may use a cross-coupled power clamp between the domains (Fig. 5.43b).

A similar approach is used in the case of a two-voltage domain with a distributed active clamp network (Fig. 5.44).

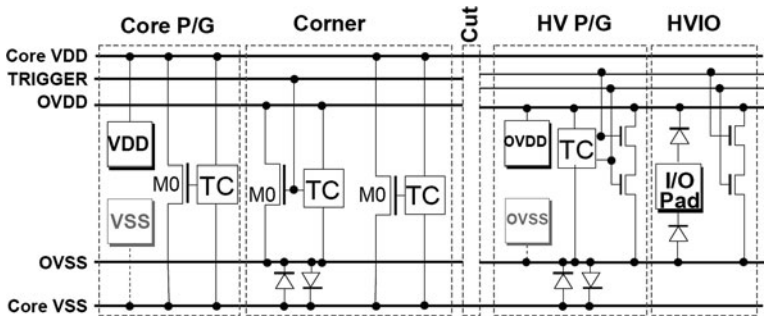


Fig. 5.44 ESD protection network for two-voltage domains with separate distributed active clamp [104]

5.3.2 Protection of Multiple Voltage Domains with Single Active Clamp Network

In small pin count, low-power analog circuits with multiple voltage domains, the protection network often needs to be implemented using only one active clamp domain. In this case, a common ground can be used for all the power supply domains. The principles for a successful implementation of this structure are summarized below.

The active clamp ESDMINUS rail is connected to the ground of both power supplies.

The ESDPLUS rail of the active clamp network is connected to the highest voltage power supply. If both supplies have the same voltage level, then the ESDPLUS rail should be connected to the power supply whose domain creates less bus noise. Bus noise is created due to a voltage drop on the power rails and bond wires inductive load as a result of change in the current consumed by the corresponding circuit blocks ($L\mathrm{d}i/\mathrm{d}t$). For example, in the case of low-power analog and large digital domains, the analog domain might create less bus voltage noise, while in the case of power analog domain with a switching function the situation might be the opposite.

The bus voltage noise can be propagated back to the inputs of the circuit due to finite parasitic capacitance of the ESD diodes and thus might impact the performance of the analog circuit. The effect of bus voltage noise should be considered even if both grounds or both power supplies are connected to the same package pin in the final package. This major consideration is related to the bond wire inductance.

The main reason why one of the power supply domains needs to be connected to ESDPLUS rail is related to the parasitic PNP structure. This structure is formed by the p^+ -region of the Nwell ESD IO diode acting as an emitter, the N-well region of the same ESD diode acting as a base connected to the ESDPLUS, and the P-substrate connected to ESDMINUS acting as a collector. When the ESDPLUS is not connected to the power supply, this parasitic PNP can be turned on and thus might produce a significant leakage path to the ground, interfering with the power

supply. As a compromise, the ESDPLUS can be connected through an additional resistor to suppress the parasitic action of the parasitic p–n–p.

Another important consideration in such active clamp protection networks is specification of power supply operation conditions. Usually, an active clamp solution is used when power supply ramping is rather slow (100 ms). Typically, starting from a 50 μs rise time, a parasitic current through the active clamp is already unacceptable for most designs. A major limitation of active clamp application is in circuits with a "power-down" specification when IO pins remain connected to an active data bus. Both fast multiple power-down events and the loading of the external device can become undesirable and unacceptable. Thus, in the case of power-down operation with an active data bus, preference should be given to snapback solutions.

When snapback solutions are used, one of the major considerations is transient latch-up. Transient latch-up can be observed in certain cases when the holding voltage of the snapback is below the power supply voltage level. For example, in both CMOS input and output as well as BJT input (Fig. 5.45), local ESD protection with a holding voltage below the power supply level will not be able to provide a latch-up current path. However, in the BJT output, for example, in the CML output (Fig. 5.45), the latch-up consideration is rather critical. Due to termination of the 50 Ω resistors (R), the snapback device with a holding voltage below power supply can be turned on and then kept in high conductivity by the power supply VDD voltage, thus creating a latch-up scenario.

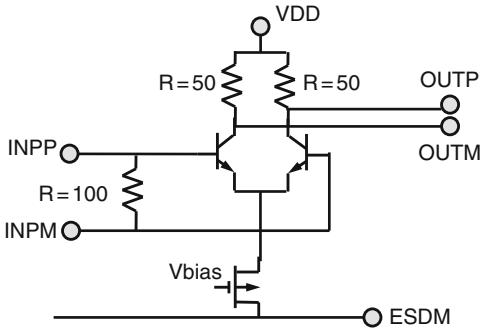


Fig. 5.45 Illustration for BJT CML inputs and outputs

5.3.3 Local Bi-directional ESD Protection of Differential Input

The case of back-drive compatible pin is one of the examples when a rail-based ESD protection network with diode clamps is not possible. The pin must be independent from power and thus the diode-to-ESDPLUS rail is not suitable.

An example of multiple voltage domains with level shifts used in LCD applications is presented in Fig. 5.46. An example of a level-shifter ESD protection network with dual-direction clamps for the display driver in presented in Fig. 5.46. This

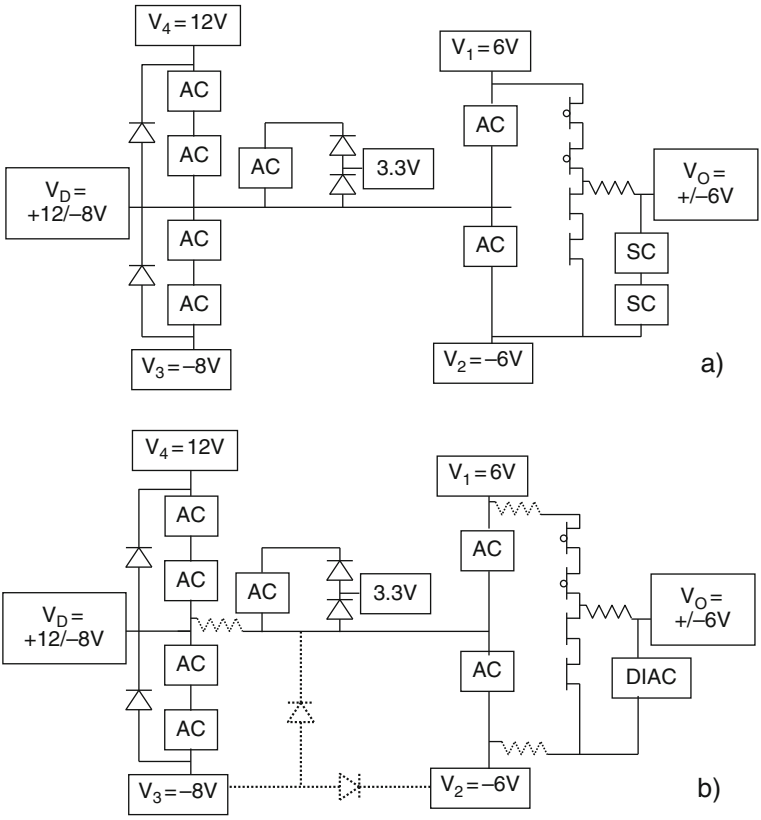


Fig. 5.46 ESD protection for LCD driver using stacked NMOS (**a**) and DIAC protection for output (**b**)

application is designed to operate at a substrate voltage of –8 V with power supply pins of –8, –6, 0, 6, and 12 V.

The analog circuit is designed using 3.3 and 6 V isolated CMOS devices. The ESD protection network combines five blocks with separate ESD clamps:

(i) Distributed active clamp (AC) blocks, with a clamp width of ~500 μm, L_g=0.8 μm, and an RC timer with RC = 4 μs
(ii) 6 V NMOS snapback clamp
(iii) Dual-direction ESD devices for 3.6 V inputs
(iv) Dual-direction ESD devices for 20 V inputs (level shifter)

The +6 V/–6 V-tolerant input pin V_O is protected by the substrate-isolated stacked snapback NMOS clamp. The isolation of the Nwell region is common for both snapback clamps to enable smaller footprint for the stacked clamp. The Nwell is connected to the +6 V power supply (Fig. 5.46a).

Alternatively, this pin can be protected by the dual-direction DIAC ESD clamp (Fig. 5.46b). The important parasitic resistance and parasitic diodes provided by the internal circuit are illustrated in Fig. 5.46b by the dotted line. The additional voltage drop and the alternative current path provided by these components of the network should be accounted for.

5.4 ESD Network Simulation with ESD Compact Models

Significant progress in predictability of ESD design can be achieved by circuit simulation using ESD compact models. The ESD compact models can be developed using standard device models and subcircuit components to model snapback regimes.

Usually, the subcircuit components include parasitic BJT structures, an avalanche current source, and saturation resistors. Development of the accurate snapback model is generally not an easy task. Due to complicated extraction methods for ESD current conditions, the construction of the scalable model requires advanced expertise in the field. One of the most helpful ways of developing a snapback model is to use the physical process and device simulation to generate the reference data for model extraction. Such an approach provides a significant acceleration of the extraction procedures while providing reasonable accuracy.

Conversely, the generated compact models can be used for both ESD and functionality simulations of the large circuit, using standard circuit simulation tools such as HSPICE and Spectre, as well as mixed-mode simulators, for example, using DECIMM^{TM} tool for this book examples.

There are several original research studies written in the field of snapback model development [107–111]. For example, the snapback ESD compact model for an LV voltage NMOS is presented in [111].

The model subcircuit uses the following components: a standard compact NMOS device model, an avalanche current source, and a parasitic NPN structure. The snapback compact model for SCR-type devices requires an additional PNP structure. With proper calibration, the model can be successfully used in simulation of ESD protection circuits.

5.4.1 Compact Model for Snapback NMOS and PMOS Devices

The cross section and the equivalent circuit representing the low-voltage NMOS device are presented in [111]. The basic cross section and equivalent circuit for the snapback model are shown in Fig. 5.47a, b, respectively.

The model combines the following components:

The standard NMOS represented by a non-snapback N-MOSFET circuit model, calibrated for a particular technology node.

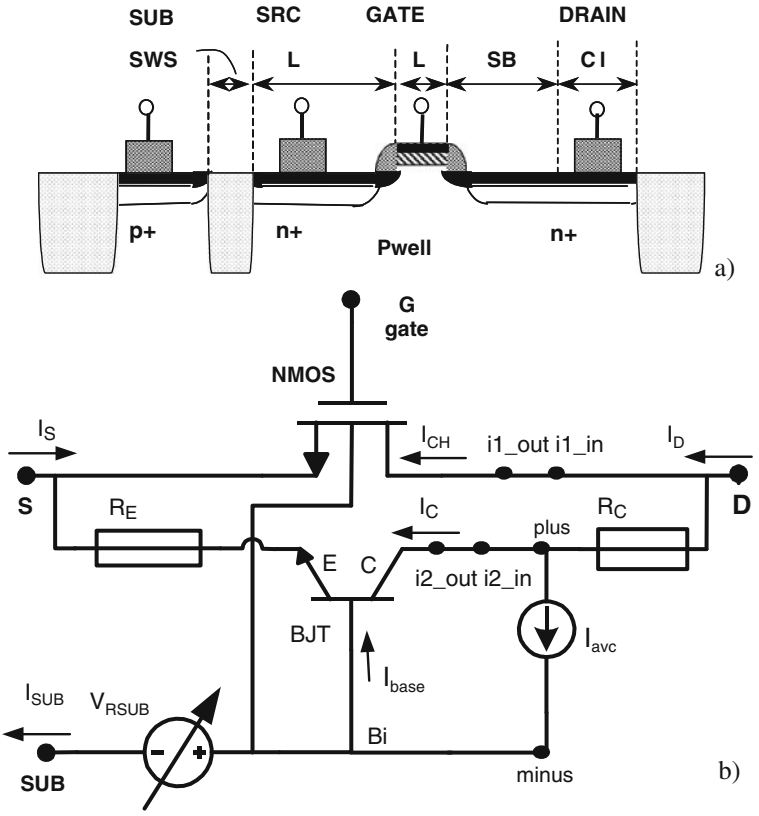


Fig. 5.47 Basic 2D cross section and ESD layout parameters of an NMOS device (**a**) and the equivalent circuit for the snapback model (**b**)

The parasitic NPN BJT represented by a basic NPN transistor model, which accounts for bipolar operation under ESD conditions with extracted parameters for current gain, forward ideality factor, reverse ideality factor, equivalent high-current collector resistor, and equivalent high-current emitter resistor.

VRSUB is the current-controlled voltage source that represents the substrate resistance modulation. It is modeled as:

$$V_{R_{\mathrm{SUB}}} = \frac{I_{\mathrm{SUB}}(L + L_{\mathrm{s}} + \mathrm{SWS})^{\mathrm{alfa\_}R\mathrm{sub}}}{a\_R\mathrm{sub} \times W - b\_R\mathrm{sub} \times I_{\mathrm{SUB}}}, \tag{4.4}$$

where L, L_s, SWS, and W are geometrical parameters of the structure indicated in Fig. 5.47a, and $a_R$sub, $b_R$sub, alfa_Rsub are extracted model parameters from the experimental data [111].

The component I_{avc} represents the avalanche breakdown current source. It is described using the following model:

$$I_{avc} = (M_b - 1)I_c + (M_{ch} - 1)I_{ch}, \tag{4.5}$$

where the current components I_c and I_{ch} are as shown in Fig. 5.47a, and M_b and M_{ch} are the avalanche multiplication factors due to the bulk and channel regions with a high electric field in the vicinity of the drain–substrate junction depletion area.

The structure of the snapback PMOS model is essentially the same as that of the NMOS model above, with the changes from NMOS to PMOS and NPN BJT to PNP BJT.

5.4.2 Snapback LVTSCR Model

By using appropriate subcircuit components to provide positive feedback on the device level, the LVTSCR model can also be created. Circuit components in this model include additional devices (Fig. 5.48) to those already in the snapback NMOS compact model.

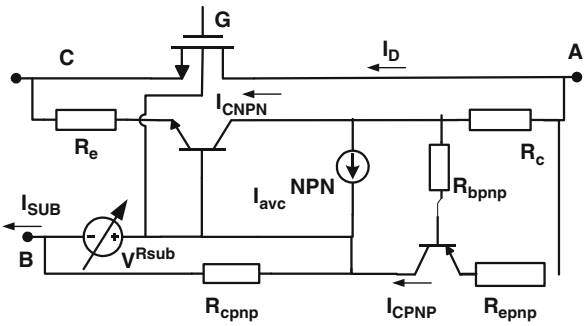


Fig. 5.48 Equivalent circuit for an LVTSCR compact model

To reproduce the silicon-controlled rectifier, the major additional component needed is the PNP bipolar transistor. Together with the parasitic NPN already present in the circuit, this circuit creates a typical Ebers equivalent of an SCR. Additional components include the resistors R_{bpnp} to control the base resistance, R_{epnp} to control the emitter resistance, and R_{cpnp} to control the collector resistance of the effective PNP device.

5.4.3 Extended Voltage Snapback Compact Models

The most advanced models that have been developed support extended voltage and high-voltage devices. Simplified cross-sections are presented below for extended

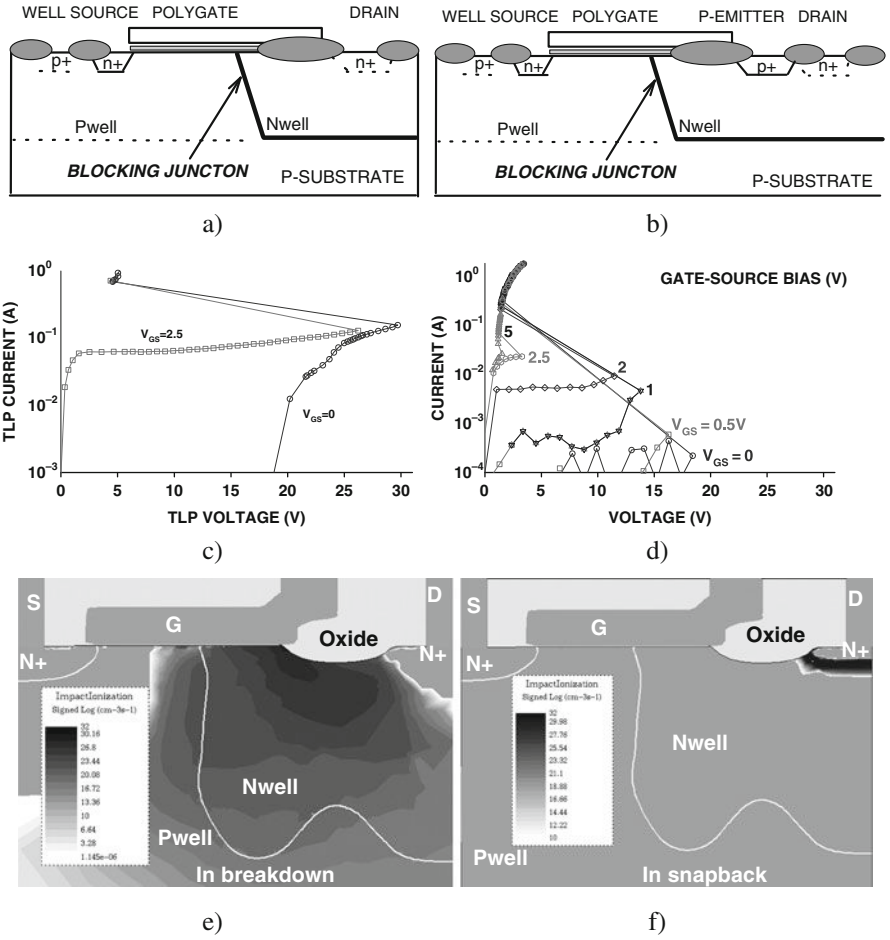


Fig. 5.49 Simplified cross section of 12 V NLDNMOS (**a**) and 12 V/20 V NLDMOS-SCR (**b**) devices and corresponding measured TLP characteristics (τ_P =100 ns, τ_R=10 ns) at different gate bias values for $W = 200$ μm (**c**, **d**). Simulation of the Kirk effect inside the NLDMOS device as a change of impact ionization region localization for the condition of avalanche breakdown (**e**) and snapback (**f**)

voltage NLDMOS (Fig. 5.49a) and NLDMOS-SCR (Fig. 5.49b) devices, with corresponding TLP characteristics shown in Fig. 5.49c, d, respectively [112].

Equivalent circuits of the snapback compact models for the NLDMOS and NLDMOS-SCR are presented in Fig. 5.50.

For the NDeMOS device at high drain–source bias, the breakdown is initiated in the depletion zone of the Nwell–Pwell junction, as illustrated in the TCAD simulation results shown in Fig. 5.49e, f. Increasing the drain bias causes a significant voltage drop across the low-doped Nwell drain region, since it is depleted of carriers. This is seen in the *I–V* characteristics after junction breakdown. When the

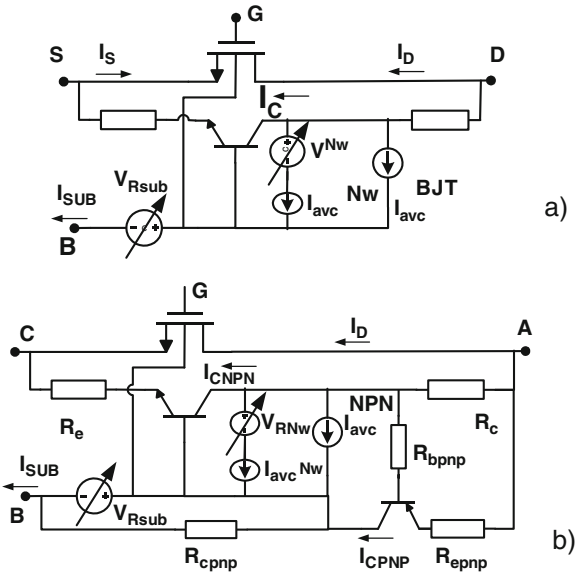


Fig. 5.50 Equivalent circuit of the snapback model circuits for NLDMOS (**a**) and NLDMOS-SCR (**b**)

high-field depletion region expands with increasing drain bias, it touches the highly doped N^+ drain region. The conductivity of the Nwell region becomes completely over-modulated by the generated electrons and holes. This results in a shift of the maximum electric field (and impact ionization generation) toward the N^+ /Nwell interface, as shown in Fig. 5.49f. Often, this effect is referred to as the Kirk or base-push out effect [14]. In this regime, the parasitic bipolar device is activated, triggering the structure into snapback.

This behavior of the device has been used to develop circuit models for NDeMOS and NDeMOS-SCR structures operating in ESD conditions [112].

Initially (Fig. 5.49a), at low drain bias, the equivalent bipolar device and avalanche current source I_{avc}^{BJT}, representing the impact ionization at the N^+ /Nwell junction, are not active. The I_{avc}^{Nw} avalanche current source describes the breakdown of the Nwell–Pwell junction and is expressed as $I_{avc}^{Nw} = M_{ch} I_S$. M_{ch} is the multiplication factor for the channel current and is given by the standard model

$$M_{ch} = \frac{1}{1 - \left(V_D / V_{BR}^{Nw}\right)^n} - 1. \tag{4.3}$$

V^{Nw} represents the voltage drop across the carrier-modulated Nwell region. It accounts for the modulation of Nwell resistance with the increase of injected carriers and is modeled as

$$V^{\mathrm{NW}} = \frac{I_{\mathrm{avc}}^{\mathrm{NW}} l_{\mathrm{W}}}{A_{\mathrm{eff}} q \mu_n \left(N_{\mathrm{d}} + \frac{I_{\mathrm{awc}}^{\mathrm{NW}}}{A_{\mathrm{eff}} q v_{\mathrm{sat}}} \right)}. \tag{4.4}$$

In (4.4), l_{w} and A_{eff} represent the effective length and cross section of current flow in the Nwell region and are treated as fitting (extracted) parameters; q, μ_n, and v_{sat} are the electron charge, mobility, and saturation velocity, respectively, while N_{d} is doping. A similar equation is used for modeling the voltage source $V^{R\mathrm{sub}}$ that represents the increase of substrate potential in junction breakdown conditions.

The avalanche current source $I_{\mathrm{avc}}^{\mathrm{BJT}}$ represents the shift of the avalanche region from the Nwell–Pwell to the N^+ /Nwell region, which leads to the activation of the parasitic bipolar structure. $I_{\mathrm{avc}}^{\mathrm{BJT}}$ is described as $I_{\mathrm{avc}}^{\mathrm{BJT}} = M_{\mathrm{BJT}}(kI_{\mathrm{S}}+I_{\mathrm{C}})$. M_{BJT} is described using a similar equation, as M_{CH} (with different breakdown parameters) and k are parameters used to control the gate-coupling effect in snapback operation. Note that when $I_{\mathrm{avc}}^{\mathrm{BJT}}$ is activated, $I_{\mathrm{avc}}^{\mathrm{Nw}}$ consistently self-deactivates due to the bias V^{Nw}. This voltage drop reduces the effective bias across the Nwell–Pwell junction and, correspondingly, the avalanche multiplication that generates $I_{\mathrm{avc}}^{\mathrm{Nw}}$. Such model behavior is equivalent to the above-mentioned shift of the avalanche region inside the Nwell.

A properly calibrated compact model can provide a rather good match between the TCAD *I–V* characteristics calibrated to the experimental data and the circuit simulation results obtained using the model [112]. The model also accurately represents the breakdown and triggering behavior for different structure dimensions.

Figure 5.51a, b shows the HBM response of the reference TCAD structure (a) and the circuit model (b) for different HBM precharge levels. Initially, the drain voltage increases with the transient increase of the amplitude of the incident HBM pulse. The peak voltage indicates the snapback triggering condition, which causes collapse of the voltage. When the discharge current becomes sufficiently low, the device switches off and the voltage starts to increase again, due to the charging of the drain capacitance by the small remaining HBM discharge current. It is seen that at low HBM precharge voltage, the structure tends to oscillate due to the very low discharge current that is not sufficient to support the bipolar snapback operation. This is captured equally well by the circuit model.

Similar to LVTSCR, described above, the NDeMOS-SCR model is constructed by attaching a PNP bipolar device to the NDeMOS snapback model, as shown in Fig. 5.50b. The equivalent emitter and base resistance R_{epnp} and R_{bpnp} are used to control the activation and the high-current operation of the equivalent PNP transistor. R_{cpnp} represents the Pwell resistance.

Simulated snapback triggering behavior of the NDeMOS-SCR device at different gate bias conditions is presented in Fig. 5.51c and correlates with the experimental results (Fig. 5.49d). The compact model reproduces the failing dependence of the triggering voltage upon the gate bias very well.

To illustrate the usefulness of the compact models, an example of a circuit simulation reproducing the snapback of the internal circuit is presented in Fig. 5.52. The original circuit failure analysis has indicated failed internal driver components.

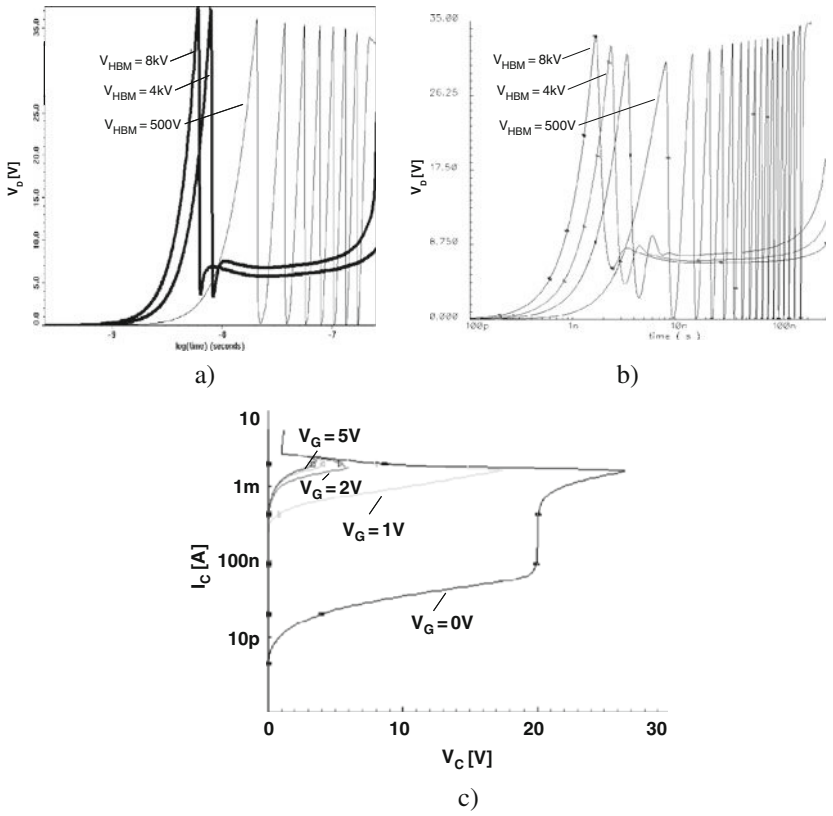


Fig. 5.51 HBM response of the reference-calibrated NDeMOS TCAD structure (**a**) and the compact model (**b**) for different HBM precharge voltages. DC snapback simulation using the NDeMOS-SCR model at different gate bias conditions (**c**)

The simulation schematic (Fig. 5.52a) is composed of the low voltage 5 V snapback NMOS M_1, which represents the gate load of the large 20 V NLDMOS power array protected by NLDMOS_SCR snapback ESD clamp. The ESD pulse for the analysis is produced by the HBM voltage source.

The circuit can be used as a representation of the switching output stage of the DC–DC converter.

According to simulation results for the waveforms of the ESD signal transmitted through the parasitic gate-drain or gate-source capacitances, the internal capacitor provides supply voltage of the driver circuit at the snapback voltage level (Fig. 5.52b). Taking into account the possible non-uniform current distribution across the array, snapback of 5 V NMOS representing the driver circuit can be expected.

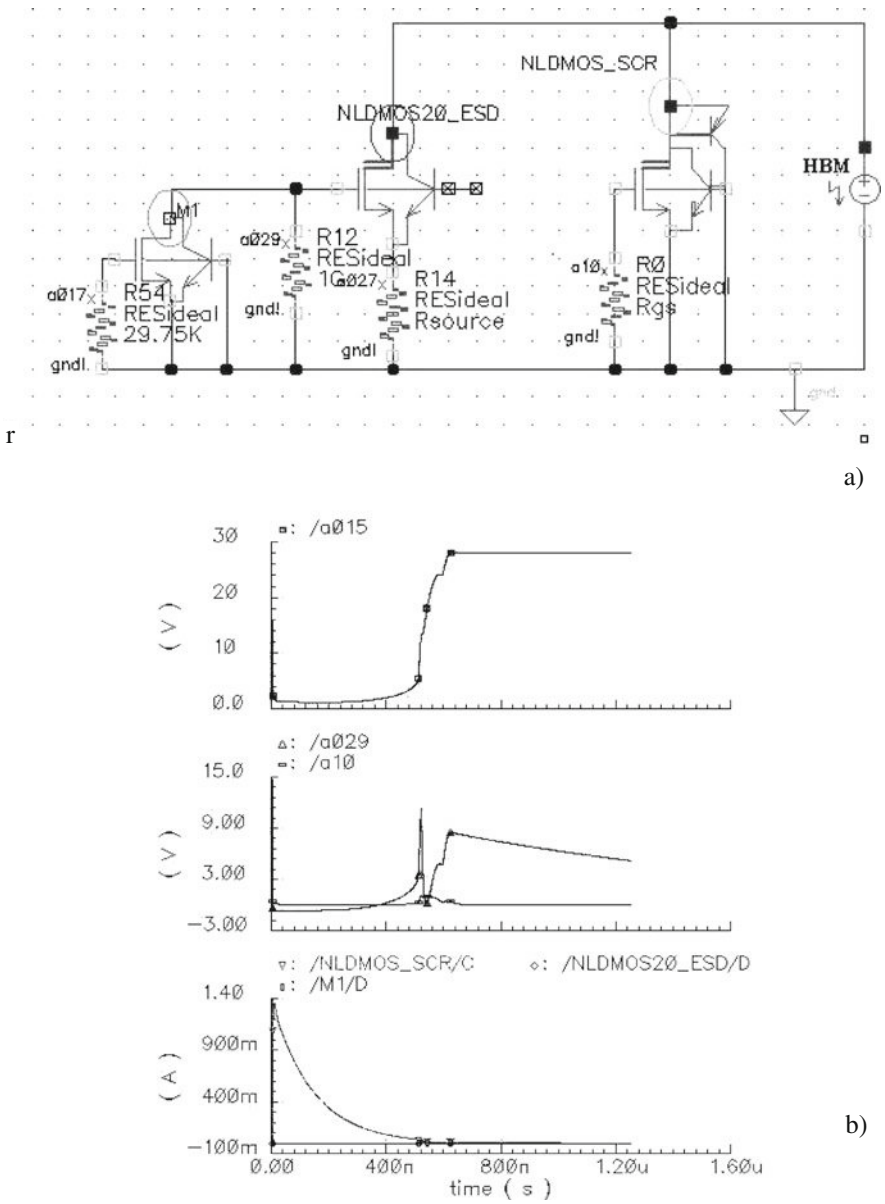


Fig. 5.52 Simulation circuit (**a**) and waveforms (**b**) for the voltage and current at the highlighted circuit nodes

5.4.4 High-Voltage Open Drain Circuit Analysis

Another example (Fig. 5.53) is presented by a simplified schematic of an open-drain driver circuit, which combines NDeMOS and NDeMOS-SCR structures. The NDeMOS device has an attached 30 kΩ gate resistor to mimic the dynamic coupling

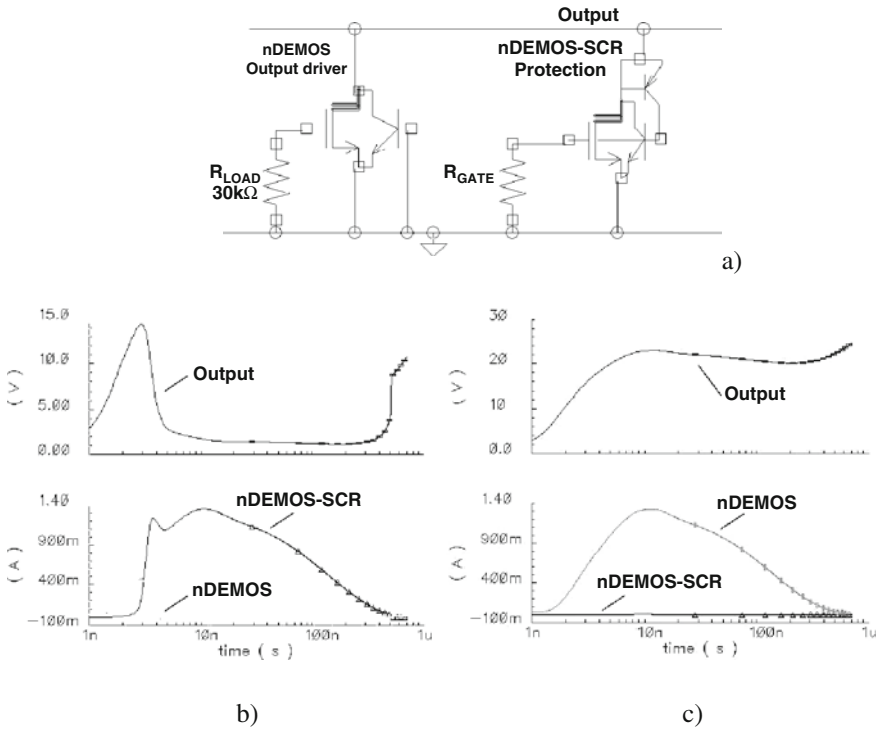


Fig. 5.53 Simplified schematic of ESD protection for open-drain output driver (**a**) voltage and current waveforms for two values of RGATE – 10 kΩ (**b**) and 1 kΩ (**c**) representing two possible circuit operation modes

load from the driver circuit. In ESD conditions, depending on the circuit and structure parameters, the ESD current can discharge either through the NDeMOS-SCR ESD protection clamp or through the NDeMOS device.

As can be seen from the comparative analysis of the waveforms, two different scenarios are observed. When $R_0 = 10$ kΩ (Fig. 5.53b), under given circuit parameters for a 2 kV HBM, the gate coupling on the NLDMOS-SCR gate is sufficient to provide the early turn-on. As a result, the SCR clamp takes over the ESD current. On the contrary, in the case of insufficient gate coupling where $R_0 = 1$ kΩ (Fig. 5.53c), the clamp does not turn on and the current path is formed through the 5 mm NLDMOS device.

Thus, depending on the circuit parameters and the HBM pulse amplitude, the critical regime can be determined as a condition of where the current changes direction in the turning-on circuit from the SCR clamp path to the NLDMOS path. If the current through NMOS is uniform, the high-current operation will be non-destructive. However, in a real 3D situation, the local snapback turn-on might result in irreversible failure [113].

A similar scenario of competition between the snapback clamp and array turn-on can be a result of the substrate potential effect. The substrate coupling technique is also one of the most useful methods in controlling ESD devices [78].

5.5 Summary

The ESD protection network essentially represents a pulsed power circuit that provides different current paths for each pin-to-pin combination. In spite of large variety of analog circuits, the network can be constructed based on common and rather simple principles. These include appropriate selection of ESD clamps, metallization routing adequate for high ESD current, and accounting for the alternative current path(s) throughout the internal circuit components.

The last is a very important design factor that cannot be neglected, but is used instead to provide appropriate performance of the circuits. In most practical cases, multiple possible scenarios of internal current paths should be analyzed, followed by experimental verification of the adequate operation.

The choice of the local or rail-based network in different power domains is application-specific, as well as the choice between the active and snapback clamps.

Since an analog circuit pin can interface with many active devices connected to the pin, both mixed-mode simulations and circuit ESD simulations with compact modeling tools can provide significant help in understanding the coupling of internal circuit components and current path formation. The same approach can be used to validate or estimate the effectiveness of the final measures.

The ability to make the right choice greatly depends upon the understanding of the ESD protection network design and the components used in it, while simply reusing existing solutions does not always prove adequate.

DECIMM^{TM} Simulation Examples for Chapter 5

To download a trial version of the numerical simulation software and request an electronic license key please visit http://www.analogesd.com

To download libraries with simulation examples for this chapter please visit http://www.analogesd.com/Chapter5.html

List of examples is subject to change.

Example 5.1 Active 5 V NMOS Clamp

Library Name: Examples5A_ESD_Active_Clamps
Project Name: E5.1a_5 V_AClamp_HNMOS
Project Name: E5.1b_5 V_AClamp_PowerUp
Project Name: E5.1c_MM_AClamp_HNMOS
Project Name: E5.1d_5 V_AClamp_Consolidated

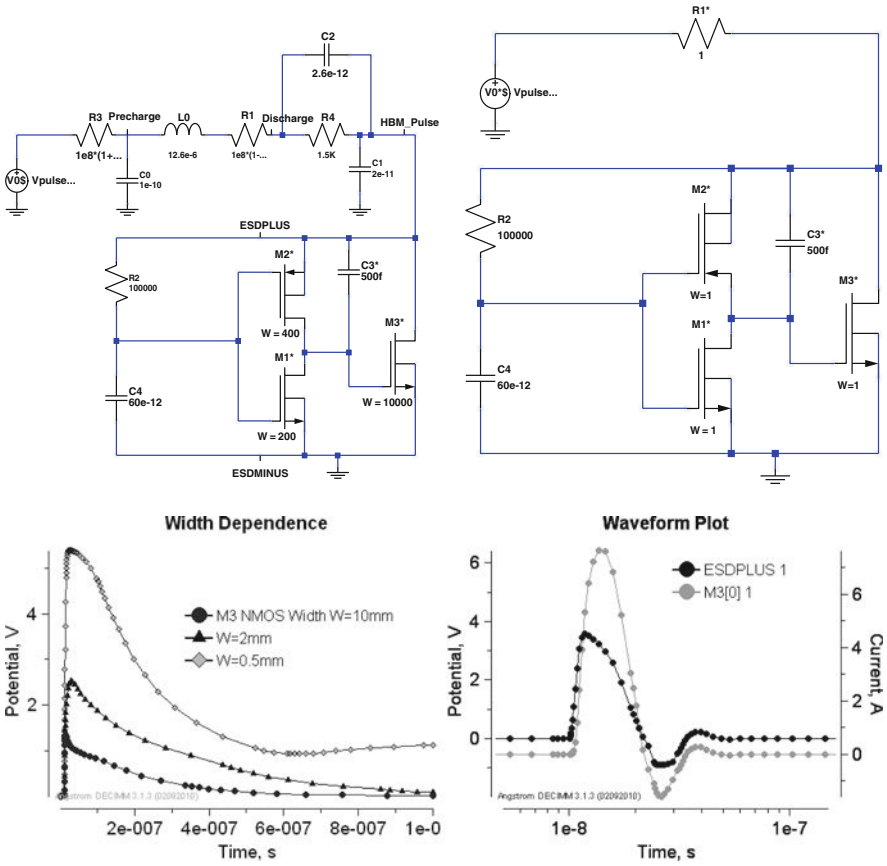


Fig. E5.1 Mixed-mode simulation circuits of the active NMOS clamp for HBM ESD, MM ESD, and power supply voltage ramp simulation and simulated circuit responses for 2 kV HBM, 200 V MM ESD pulses

This set of examples is prepared for exploration of the most popular active clamp solution based upon the NMOS device. The first example demonstrates waveforms produced by the clamp under an HBM ESD pulse and can be used to study the effect and optimization of the RC timer and driver parameters as well as the size of the NMOS (Fig. E5.1). The second example shows the dependence of the parasitic current of the clamp for the power supply ramp. The following example provides the mixed-mode circuit for clamp operation under the machine model (MM) ESD pulse. The last example has all three circuits consolidated in a single project.

Example 5.2 Active 5 V PMOS Clamp

Library Name: Examples5A_ESD_Active_Clamps
Project Name: E5.2_5 V_AClamp_HPMOS

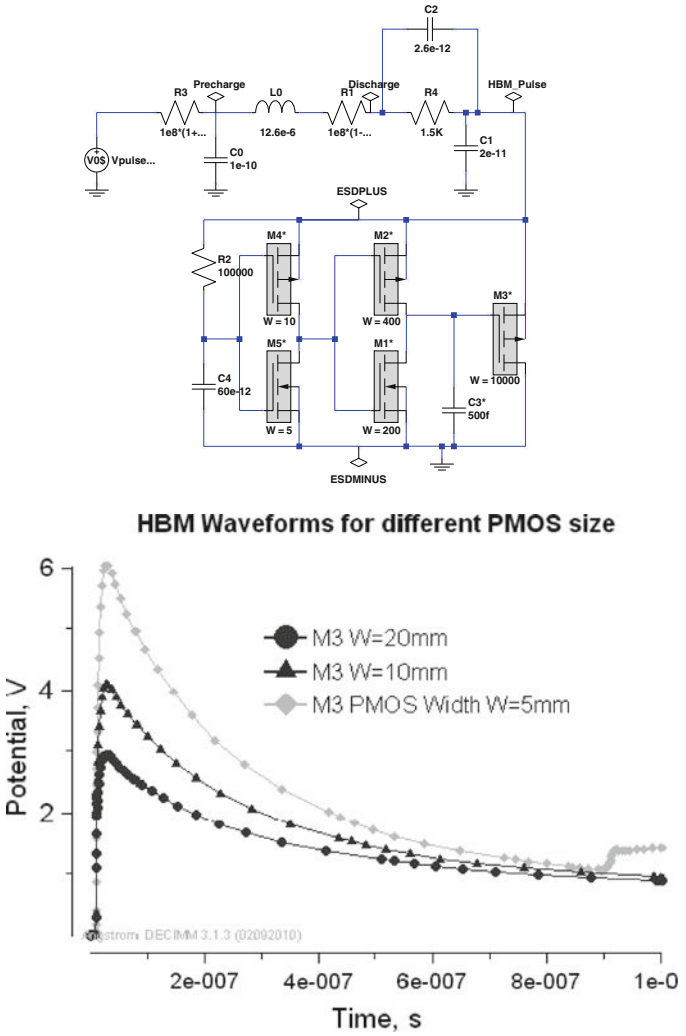


Fig. E5.2 Mixed-mode simulation circuit for the active PMOS clamp and simulated waveforms under HBM ESD pulse stress

This example presents the 5 V PMOS-based active clamp (Fig. E5.2). In some cases, this type of clamp is preferable over the NMOS clamp due to the high holding voltage of the PMOS and corresponding elimination of the transient latch-up conditions.

Example 5.3 EEPROM Erase pin Protection

Library Name: Examples5A_ESD_Active_Clamps
Project Name: E5.3a_EEPROM_Erase_Clamp_ver1
Project Name: E5.3a_EEPROM_Erase_Clamp_ver2

In this example, two alternative versions of the high-voltage erase pin protection design are compared. Both clamps rely on the high-voltage NLDMOS-SCR snapback device (Fig. E5.3). High residual voltage is built up in the first clamp, providing undesired overstress of the pin that may erase information in the memory. The second version of the clamp utilizes additional NLDMOS-based RC-controlled circuitry to eliminate the high voltage peak after SCR turn-off.

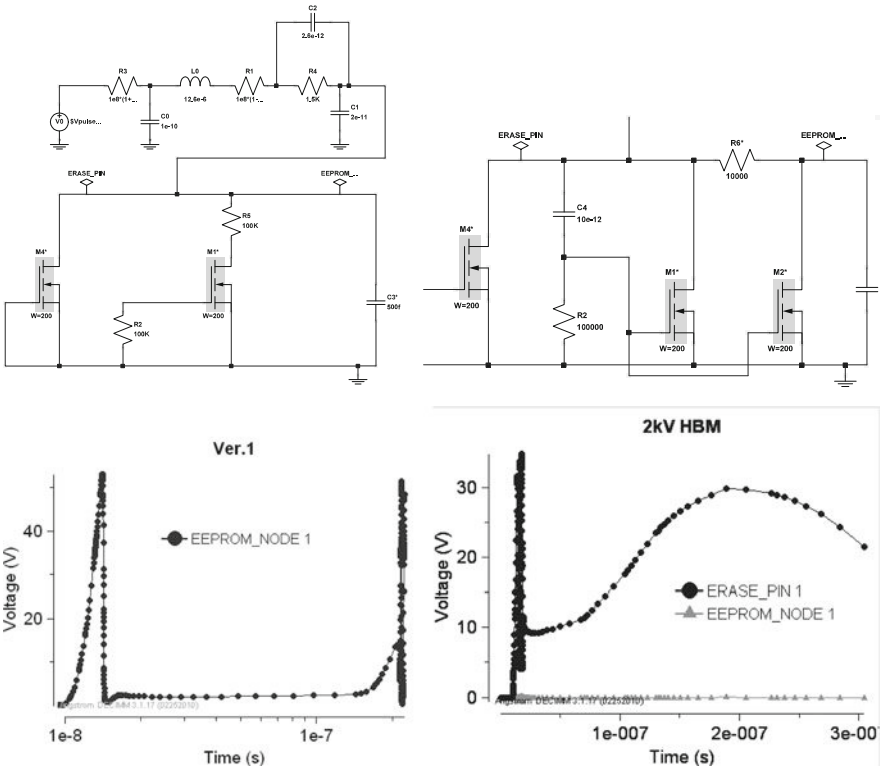


Fig. E5.3 Mixed-mode simulation circuits for the two versions of the erase pin protection clamp design with the simulation results for HBM ESD pulse conditions

Example 5.4 BJT-Based Active Clamps

Library Name: Examples5A_ESD_Active_Clamps
Project Name: E5.4a_5 V AClamp_NPN
Project Name: E5.4b_5 V AClamp_NPN_casc
Project Name: E5.4b_5 V AClamp_PNP

This set of examples presents active clamps with the BJT device as an active component. BJT-based active clamps provide a significant advantage in clamp size. Three alternative versions of the BJT-based active clamp designs cover NPN, cascaded

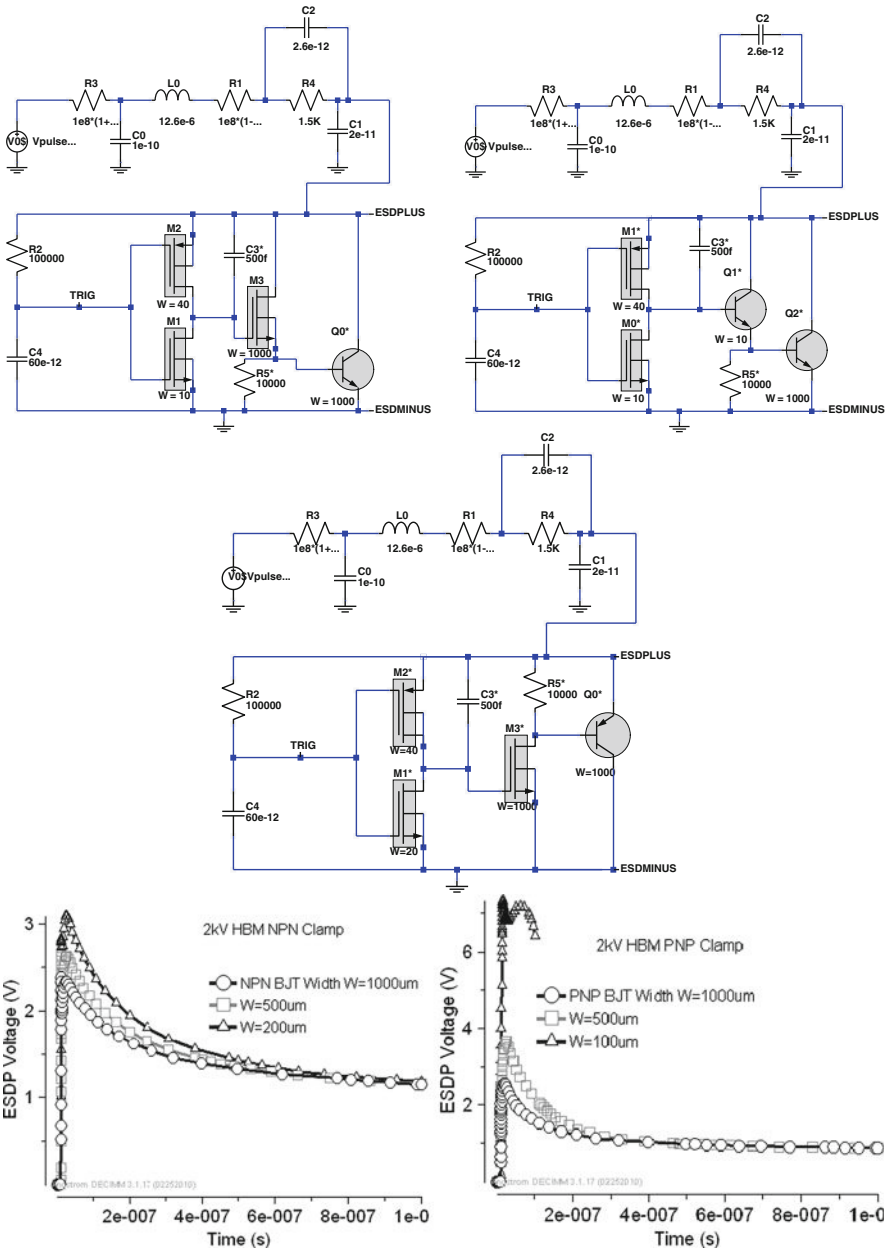


Fig. E5.4 Simulation circuits for three versions of the BJT-based active clamps and the results of the 2 kV HBM ESD pulse mixed-mode analysis for different clamp parameters

NPN and PNP versions. The simulated waveforms for these clamps can be compared with the CMOS versions (Fig. E5.4). These examples can be used as a starting point for design and optimization of the CMOS driver and BJT clamp device.

Example 5.5 Stacked Active Clamps for High Voltage Tolerance

Library Name: Examples5B_Stacked_Active_Clamps
Project Name: E5.5_Stacked_AClamp_CMOS_Consolidated

This set of examples presents an active clamp solution that achieves a voltage tolerance two times higher than that of non-stacked active clamps. Two examples present both ESD operation of the clamp and power supply ramp transient analysis. The examples are built using compact models for a 1.5 V CMOS process with only 1.5 V poly-capacitors available (Fig. E5.5). The circuit is tolerant of 3 V interface voltage. A simulation with a power supply voltage source demonstrates the current level consumed by the clamp at rise times of 10 and 100 μs.

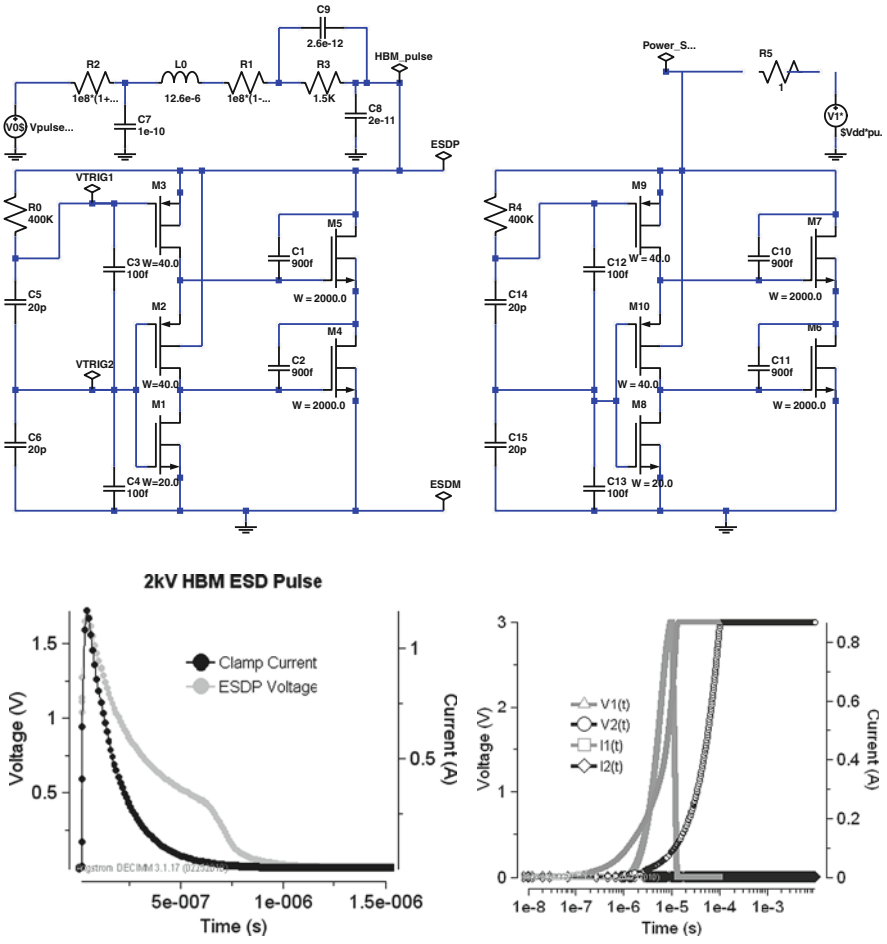


Fig. E5.5 Simulation results for the stacked CMOS active clamp with 2 kV HBM ESD pulse and power supply ramp sources

Example 5.6 Stacked Active Clamps with NPN

Library Name: Examples5B_Stacked_Active_Clamps
Project Name: E5.6a_Stacked_AClamp_NPN_HBM
Project Name: E5.6b_Stacked_AClamp_NPN_PowerUp

This and the following examples demonstrate the non-snapback ESD solution implementation toward a smaller footprint by taking advantage of BJT devices. The circuit represents the clamp assembled using a stacked driver and built with compact models for 1.5 V CMOS components combined with a FEM NPN BJT device. The circuit is tolerant of a 3 V voltage. Using the simulation, optimal driver parameters can be found for matching the driver circuit with the FEM BJT (Fig. E5.6).

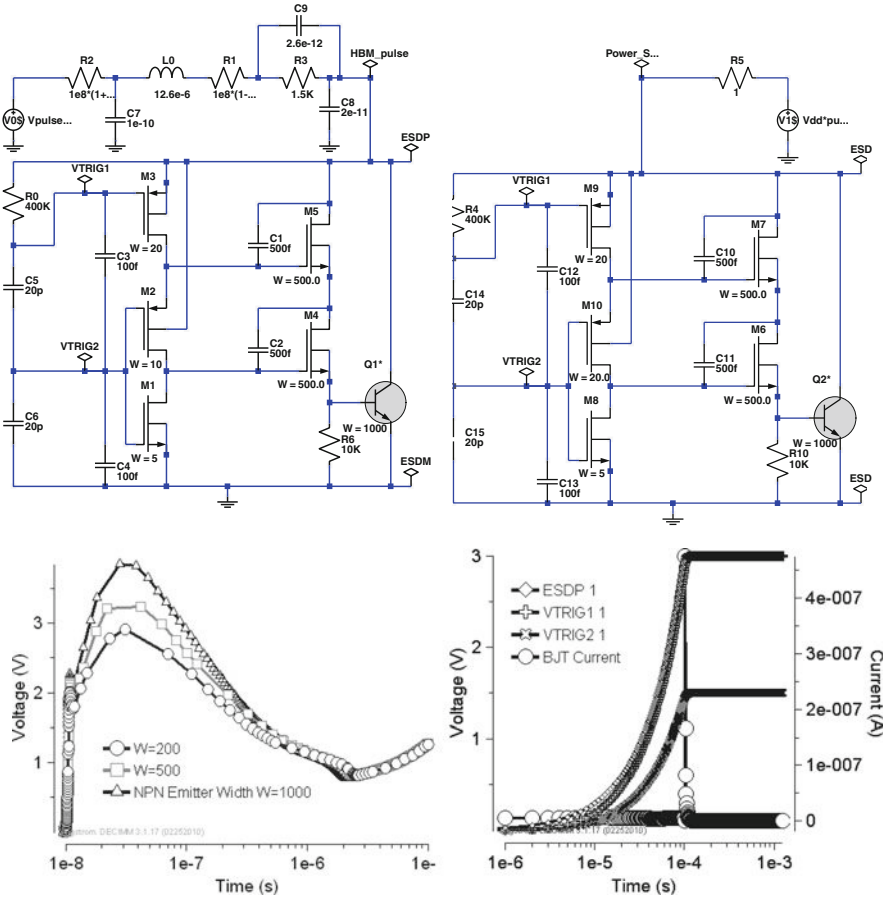


Fig. E5.6 Mixed-mode simulation circuits for the NPN-based active clamp with a stacked CMOS active clamp driver and examples of 2 kV HBM ESD pulse and power supply ramp waveforms observed in each circuit

Example 5.7 Stacked Active Clamps with PNP

Library Name: Examples5B_Stacked_Active_Clamps
Project Name: E5.7a_Stacked_AClamp_PNP_HBM
Project Name: E5.7b_Stacked_AClamp_PNP_PowerUp

This example demonstrates ESD solution implementation using PNP BJT devices. This approach is useful in a number of BiCMOS process technologies with low-voltage CMOS. A mixed-mode simulation circuit represents the clamp, which is assembled using a stacked driver built using compact models for 1.5 V CMOS components combined with a FEM NPN BJT device. The circuit is tolerant of 3 V voltage. Using this simulation, driver parameters optimal for matching the driver circuit with the FEM BJT can be found (Fig. E5.7).

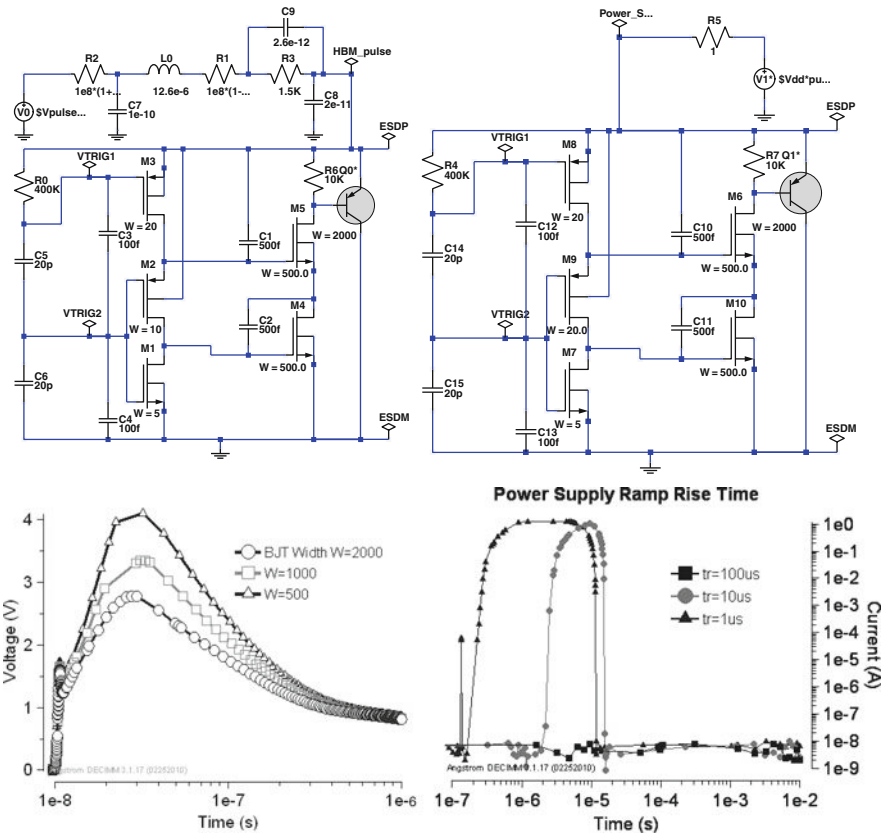


Fig. E5.7 Mixed-mode simulation circuits for PNP-based active clamp with stacked CMOS active clamp driver and examples of 2 kV HBM ESD pulse and power supply ramp waveforms observed in each circuit

Chapter 6
ESD Design for Signal Path Analog

The purpose of this and the following chapters is to demonstrate the implementation of different ESD protection approaches specific to analog products. Two major categories of analog products are used as examples demonstrating ESD protection challenges and solutions: the signal path and the power management products. The first category is addressed in this chapter. Chapter 7 is focused on the specifics of power management products. These products include low-voltage and high-voltage integrated DC–DC converters and controllers, LED and display drivers, and other power products.

In this chapter, ESD protection implementation is presented for signal path products, represented by high speed, precision and audio amplifiers, digital-to-analog converters, and interface circuit blocks.

The subdivision of analog products into signal path and power management analog circuits emphasizes a corresponding difference (based on pin specifications) in the implemented ESD protection for the two cases. While ESD protection solutions for control, digital, and low-power analog input pins can often be based on common principles, a number of representative, product-specific pins may require significant changes in both the clamp selection and ESD network design.

In the case of low-voltage signal path circuits, the major challenges are in the minimization of the parasitic influence of ESD protection components on the signal.

The majority of digital–analog converters and interface circuits are based on low-voltage CMOS devices with an operating voltage that corresponds to the supported DGO CMOS processes. Thus, I/O pin protection is designed for a voltage range below 1.2–5 V. The most typical exceptions are, perhaps, the external EEPROM program pin (~6 V) and the erase pins (8–18 V). Thus, corresponding solutions mainly focus on overcoming signal integrity challenges.

Unlike in interface and digital-to-analog converters, the voltage range of amplifiers is rather broad and can cover a 3–100 V operational voltage range. At the same time, even high-voltage ESD protection of amplifiers is significantly different from that of power management circuits (Chapter 7). Usually, the protection of high-voltage amplifiers can be implemented using a conventional rail-based approach with a voltage power clamp. This is possible because the requirements of high quality and linearity of the signal path output device parameters almost

V.A. Vashchenko, A. Shibkov, *ESD Design for Analog Circuits*,
DOI 10.1007/978-1-4419-6565-3_6, © Springer Science+Business Media, LLC 2010

automatically provide relatively large SOA margins for high-voltage output devices. In other words, a relatively wide ESD protection window is usually expected.

In opposite, in the case of power management circuits, the major challenges are related to the protection of the high-speed and high-voltage transient pins connected to the internal power devices. In the case of power devices, SOA limits can be very close to the absolute maximum voltage of power analog products. This creates a very narrow and sometimes even negative ESD protection window which demands rather aggressive ESD protection solutions.

Another specific of signal path analog products is related to high pin count packages. This specific is due to the fact that the charge device model (CDM) ESD pulse provides rather high ESD current conditions for the large form factor of high pin count packages, unlike in small pin count power circuits. This specific is similar to the one of high pin count digital products.

The overall ultimate goal of the remaining chapters is to demonstrate how the accumulated knowledge of Chapters 2, 3, 4, and 5, based on the physical understanding of conductivity modulation in semiconductor structures, ESD device and clamps design principles, and pulsed SOA and ESD protection network design principles, can be applied in practical analog design.

This goal is attained by means of a high-level review of the most common features of the representative analog products and ESD protection examples selected for each chapter. Each section contains a discussion of the selected, most typical ESD protection network for the given product class, and the related problems and solutions for the specific product pins based on corresponding case studies.

6.1 Amplifiers

6.1.1 Amplifier Product Families and Specifications

For ESD protection implementation, it is convenient to subdivide amplifiers into three major groups:

(i) High-speed amplifiers
(ii) Precision amplifiers
(iii) Audio amplifiers

The first group of amplifiers deals with the amplification of high-speed, low-power signals of up to 10 GHz, with corresponding figures of merit such as signal quality, noise, and the amplification factor.

The second group usually deals with a much slower signal that requires much higher accuracy and matching and stability of characteristics.

Finally, audio amplifiers handle amplification of the signal in the audio frequency range.

The application specifics imply certain peculiarities of the ESD protection design. High-speed amplifiers are mainly designed using corresponding high-speed, low-voltage processes. In this case, additional parasitic capacitance from the input ESD clamp may significantly impact the product performance.

Precision amplifiers, on the other hand, may tolerate regular ESD protection components but may be rather sensitive to the matching of the differential input component parameters. The same is true for power audio amplifiers, which can be implemented in a rather high-voltage process to achieve the appropriate output power.

Due to requirements of precision and linearity of the high-voltage output circuit components, the operation regimes for power amplifiers are located far from the physical pulsed SOA limits. Thus, high-voltage amplifiers are designed using components that usually provide a suitable ESD protection window for implementation of a rail-based network with a high-voltage ESD power clamp.

High-speed amplifiers are often used as parts of systems where a significant rate of information is transmitted by the signal. For example, high-speed amplifiers are designed to support the following applications: broadcast of video; clock generators and sync separators; security in a closed-circuit television cross-point; medical imaging in ultrasound analog front-end (AFE), base station receiver AFE's; digital-variable-gain amplifier (DVGA); mixers; low-noise amplifiers (LNA); driver receivers; pro-video cross-points; low-noise analog clock buffers; comparators; short haul optical links; power amplifier loop control; base station transmitters for RF-DVGA; and many others.

Modern trends in audio amplifier design include the increase of the power supply voltage from +/– 5 V to well over +/– 20 V. This high voltage is required for an improved performance, for example, in the signal-to-noise ratio (SNR). Another trend is the migration from Class AB to Class D.

The important factors in amplifier design are system performance and simplification of system design including reduction of the number of required external components.

To address these needs, new integrated semiconductor process technologies target several challenging process development areas. Table 6.1 demonstrates the large variety of BiCMOS semiconductor process technologies involved in supporting audio amplifier products. High-voltage processes are designed to provide a high gain at high voltage. Wide bandwidth is targeted by the development of a fully complementary process technology for a high fidelity audio performance with transistor matching and minimized thermal effects (in order to reduce or eliminate the required heat sink).

Another focus is the reduction of customer design time; the driver stage is usually the most time intensive for design. To achieve this high output power, multiple bipolar transistors can be paralleled, targeting up to ~300 W output power per channel with NPN/PNP output pairs.

Typical packaging options include a high variety of small pin count package form factors (Fig. 6.1).

Table 6.1 The large variety of BiCMOS semiconductor process technologies in the audio power amplifier products

Product family	BiCMOS process features
Class AB power amps 2 × 20/40 W	Thin EPI 50 V
Class AB power amps 1 × 80 W	Thin EPI 120 V
Class AB power amps 2 × 50/100 W	100 V DMOS
Class D power amps 6 × 50/250 W; Class D output stages 80–250 W H Bridges; Class D Controllers; eight channels + DSP	1.5 μm DMOS
±15 V analog uPot buffered; 128 step, 0.5 dB	36 V CMOS/BIP/Prec R
±5 V analog uPot buffered; 128 step, 0.5 dB; +5 V analog potentiometers – CMOS7; 128 step, 0.5 dB	12 V Bip/ Prec R
Low noise preamplifiers; digitally controlled preamps	36 V BJT/Precision R; 12 V BJT/ Prec R
Class AB power amps audio op amps; low-noise audio op amps <2.5 nV/rt Hz	Thin EPI

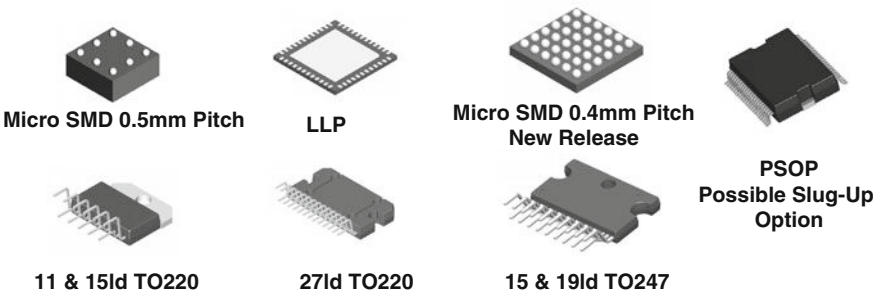


Fig. 6.1 Typical packaging options for audio amplifiers

A simplified block diagram for an audio power amplifier is presented in Fig. 6.2. The applications include high fidelity power amplifiers; high fidelity multimedia; high-performance professional audio; high fidelity equalization and crossover networks; high-performance line drivers and receivers; high fidelity active filters; very high-voltage operation; scalable output power; and feature minimum external components; external compensation; thermal shutdown; and mute.

One of the growing process-specific applications is the complementary bipolar (CB) design [114]. For broadband and narrowband, for example, wireless, the NPN/PNP symmetry is important and special attention should be paid to the thermal self-heating issues, especially in SOI process technologies (Table 6.2).

An example of the fully complementary BJT amplifier design is presented in Fig. 6.3 [114].

Examples of the specific requirements for a communication variable gain amplifier (Fig. 6.4) may include a wide band above 150 MHz, which is required from the

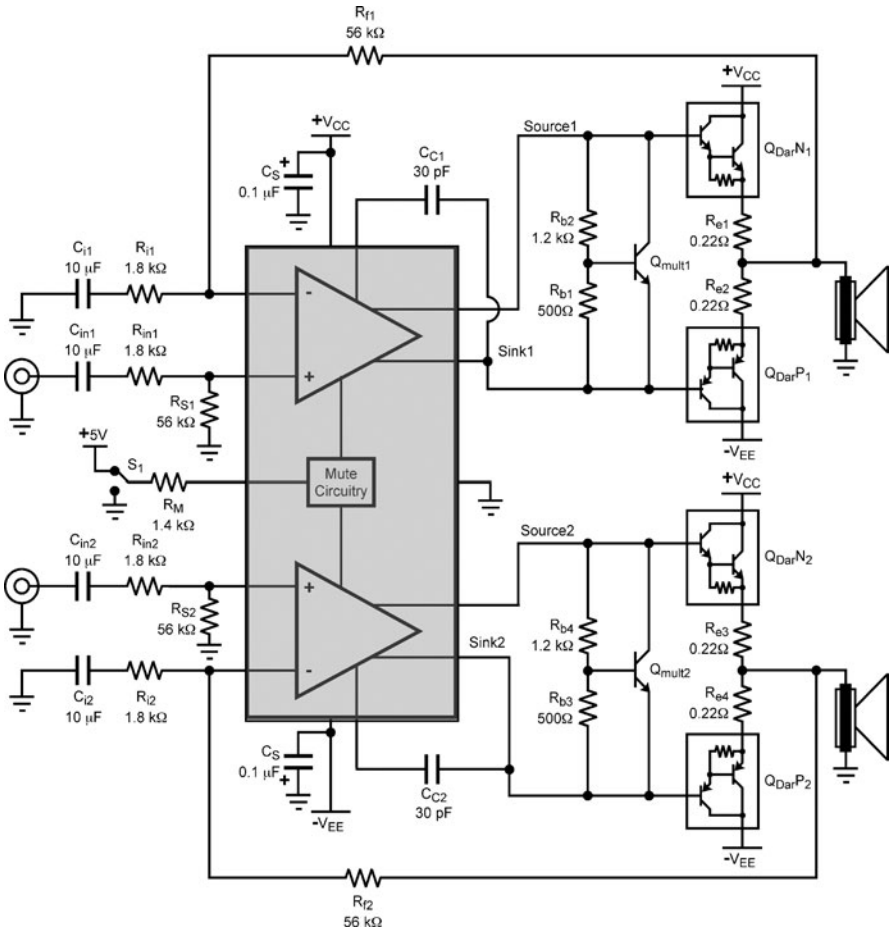


Fig. 6.2 Simplified block diagram for 2-channel audio power amplifier

process components; high f_τ and f_{max} parameters; a wide dynamic range above 80 dB provided by the logarithmic characteristics and gain matching; low voltage and current noise E_n and I_n; a high output voltage swing and drive provided by device's DC symmetry; linearity achieved by the AC symmetry.

In the case of a high-resolution video application (Fig. 6.5a), the requirements may include exceptional linearity at ~4 MHz; level of 2nd/3rd harmonics below 100 dB; 700 MHz bandwidth and 3000 V/μs slew rate. This requires high f_τ and f_{max}; exceptional gain and phase stability over the input common mode (CM) and output voltage range; 0.02% differential gain error; 0.005% differential phase error which requires AC/DC symmetry from process components; low noise of 158 dBm (1 Hz).

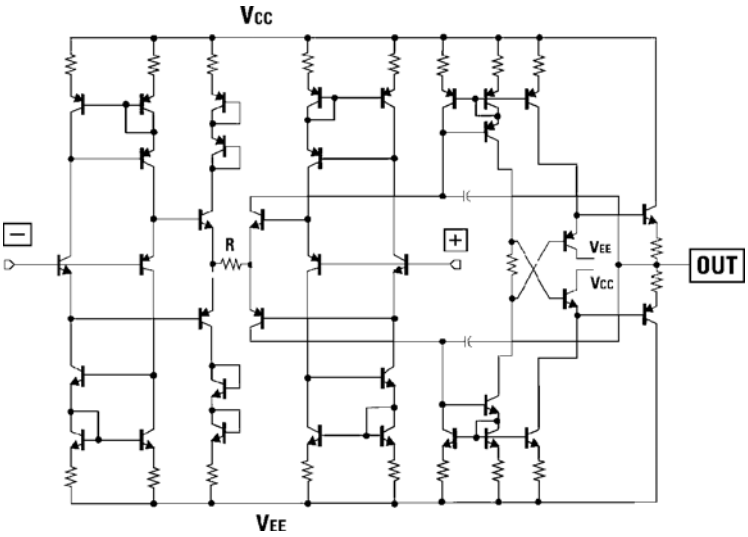


Fig. 6.3 Complementary BJT amplifier design [114]

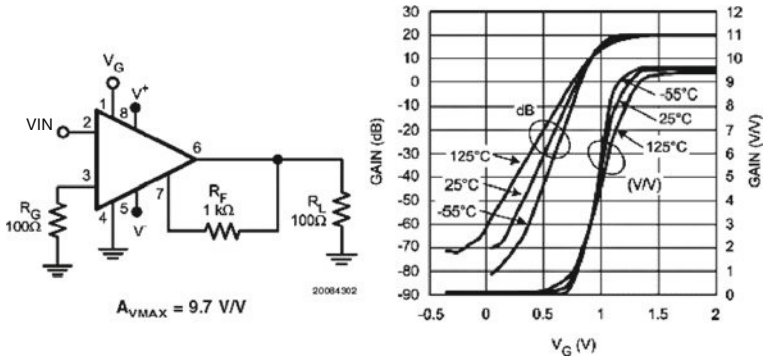


Fig. 6.4 Variable gain amp (VGA) (**a**) and dependence of the gain on gate bias voltage V_G for various temperature levels (**b**) [114]

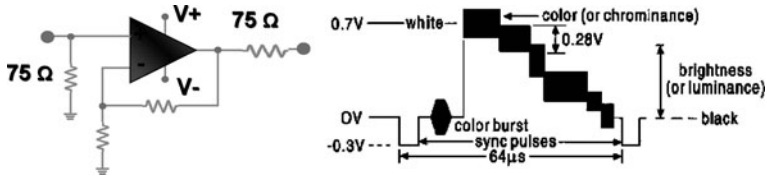


Fig. 6.5 Composite wide band amplifier and output signal at 4.43 MHz

Table 6.2 Circuit need vs. device requirements for amplifier products

High-speed amplification	High f_τ/f_{max} for NPN/PNP
Wide bandwidth at low I_{supply}	Minimize f_τ/f_{max} rolloff; minimize capacitive parasitics; oxide isolation and thin film or Poly2 resistors. Availability of MIM or Poly2-metal caps
High linearity (low distortion)	Complementary NPN/PNP BJT for f_τ/f_{max}, C_{bc}, C_{be}, C_{cs}, r_e, r_b, β, V_A
Output drive (sink/source)	Complementary NPN/PNP high current β
Low noise	Low NPN/PNP base resistance, minimize 1/*f*

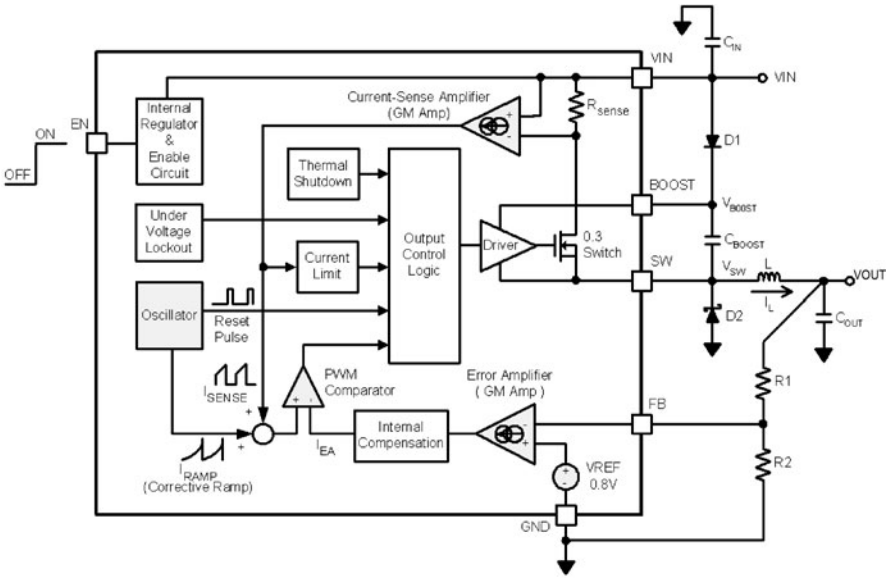


Fig. 6.6 An example of an integrated switching regulator similar to those built into power amplifier IC based upon BCD process

For portable applications, the amplifier circuit can be rather complex. To take advantage of NLDMOS devices in a high-voltage BCD process, the amplifier product may include an integrated switching voltage regulator (Fig. 6.6).

Another level of complexity can arise due to the integrated power blocks for example EEPROM charge pumps. An example of the functional block diagram for the high-speed amplifier is presented in Fig. 6.7.

The parameters of the amplifier imply a corresponding limitation on the parasitics introduced by ESD clamps and the ESD protection network. The problem cannot always be solved in a simple way. For example, the LNA for RF applications usually have a reduced ESD protection level for input pins. Due to the absence of the ESD

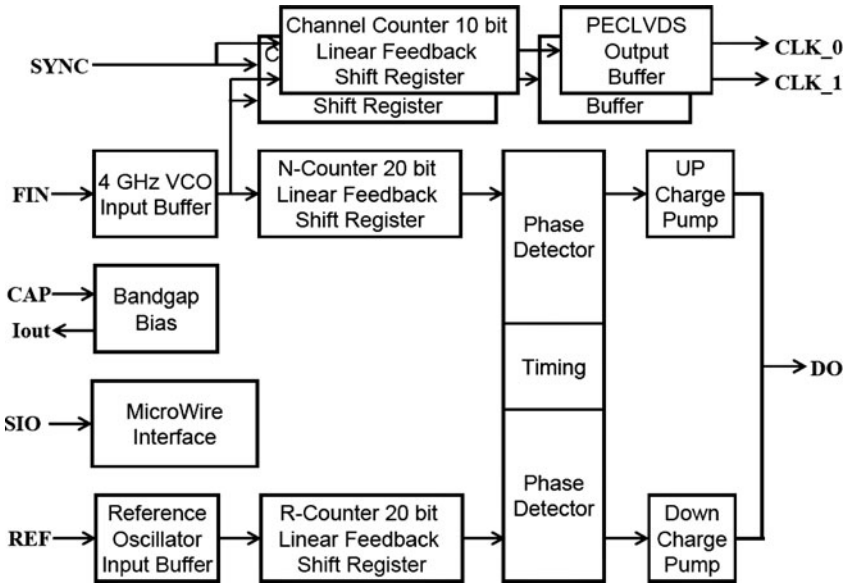


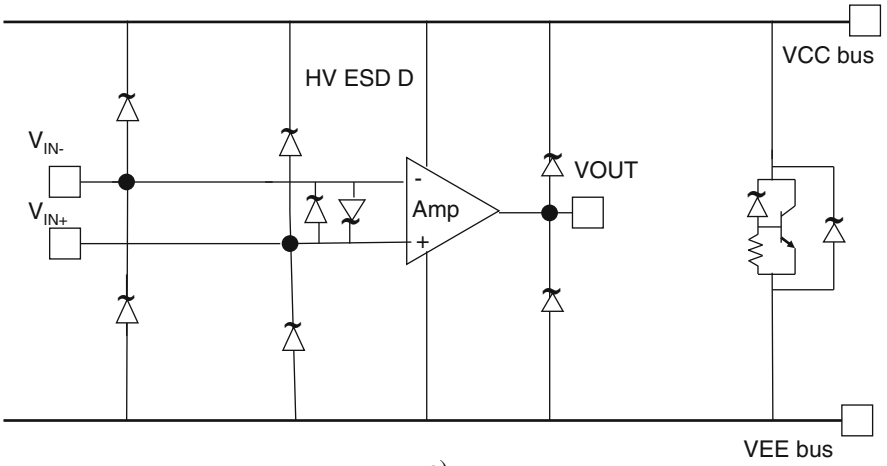
Fig. 6.7 Example of the simplified phase lock loop high-speed amplifier block diagram: phase lock loop, 5 V charge pump, programmable output buffer type, and new counter architectures with synchronization, programmable delay, and duty cycle

protection clamps, these devices rely on the internal circuit to provide at least a minimal level of ESD protection.

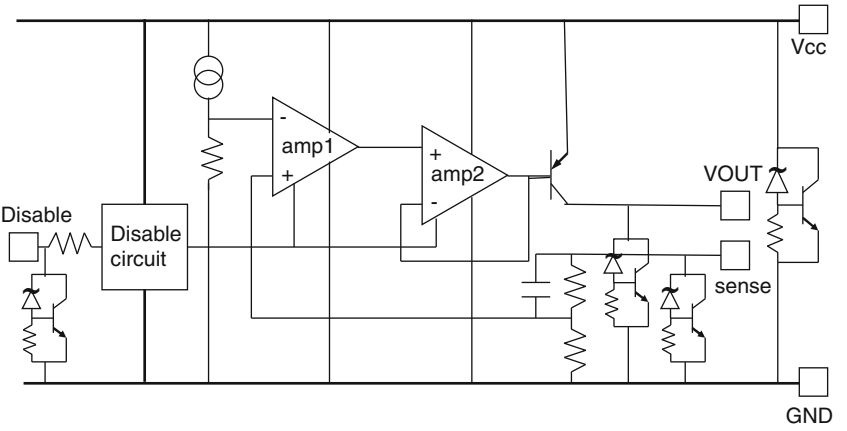
6.1.2 ESD Solutions for Amplifiers

In spite of a high variety of amplifier products with the complex integrated analog blocks described above (Figs. 6.6 and 6.7), in most cases, the ESD protection network can be successfully implemented based upon either a rail-based protection network (Fig. 6.8a) or a local clamp network (Fig. 6.8b). Such networks create an ESD current path between the amplifier pins. These pins can usually be classified as follows.

(i) CMOS or bipolar inputs, completed using a buried channel NMOS, JFET, super beta BJT devices. As an option, a differential input can be implemented.
(ii) The bipolar or CMOS outputs
(iii) A high-voltage or low-voltage power supply pins in cases of single or dual power sources. In the case of separate voltage domains, the power supply may combine both low-voltage domains for pre-amplification and high-voltage domains for power output.



a)



b)

Fig. 6.8 Block diagram for the rail-based (**a**) and local clamp (**b**) ESD protection for amplifier circuits

(iv) The additional group of pins includes low-voltage control, service, and digital interface pins.

Apparently, the amplified signal quality, as well as the performance of the whole amplifier, is the most impacted by the ESD network components connected to the output and, especially, to the input pins.

The ESD protection network for an amplifier with a single voltage domain can be realized based on a rail-based or local ESD protection network according to the

major principles presented in Chapter 5. Usually, to achieve an appropriate performance, specifically the rail-based ESD protection network is used, with I/O diodes to ESDPLUS and ESDMINUS rails. The power clamp is used to provide a current path between the ESD busses and the I/O pins.

In most practical cases, the ESDPLUS bus is not separate, but shares a rail with the power supply bus of VCC, VDD, or similar. The ESDMINUS bus is represented by a corresponding VEE, GND, or similar bus. In the case of dual voltage supply, VCC+ and VCC– can provide functions for ESD busses with substrate-isolated ESD power clamps referenced to the lower power supply node.

Typically, the snapback clamp has much bigger footprint in comparison with the ESD diodes. Therefore, a more space-saving solution is based on the rail approach (Fig. 6.8a). An additional advantage of signal integrity is provided by the lower parasitic capacitance and leakage of the diodes.

It is important to remember that although the diodes do not conduct any ESD current in the reverse-bias avalanche breakdown mode, they should still be tolerant to the corresponding high-voltage level of the given amplifier power supply domain. The same is critical for other components, for example, diffusion resistors.

Because of design constrains, it is not always preferable to provide wide metal busses. For example, in processes with fewer supported metal layers, the design of a low-resistive current path can be rather complicated and may be impacted by the signal to bus coupling limitations. Therefore, high-voltage bipolar output protection may become marginal even if the amplifier has high linearity components.

For low-cost analog process technologies with fewer layers of metal, the creation of low-impedance ESD rails across the whole chip may be sacrificed in favor of a local protection approach (Fig. 6.8b).

The product specifics of BJT or CMOS input and output designs should be considered separately and dictate the choice of the ESD protection strategy applied to the pin.

In case of hot plug-in, system-level or automotive applications additional system-level design requirements should be met for certain pins.

6.1.3 Bipolar Output High-Voltage Audio Amplifiers

Usually, it is a relatively easy task to implement a high-voltage power clamp for protection of the power supply pin only. In relatively slow high-voltage amplifiers, the design parameters for the output signal usually automatically provide a sufficient ESD protection window.

In the case of a high-voltage audio amplifier with multiple power domains (Fig. 6.9), for example, an audio amplifier with digital volume control (Fig. 6.10), diode performance becomes critical. In this case, the voltage drop on the diode should be accurately accounted for in each ESD current path scenario to ensure the absence of overstress at the I/O pins above the absolute maximum limits. For

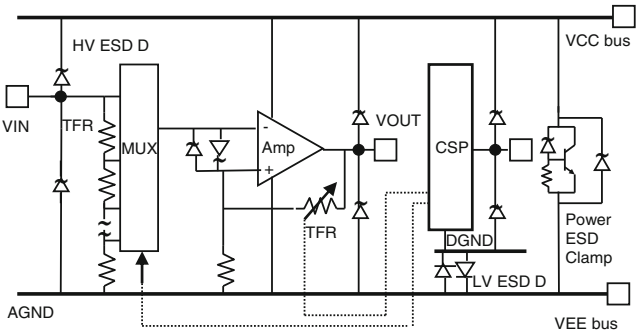


Fig. 6.9 Block diagram for two-domain ESD protection implementation in an audio operational amplifier with volume control; differential HV input, HV digital I/Os and high-voltage NPN and PNP outputs

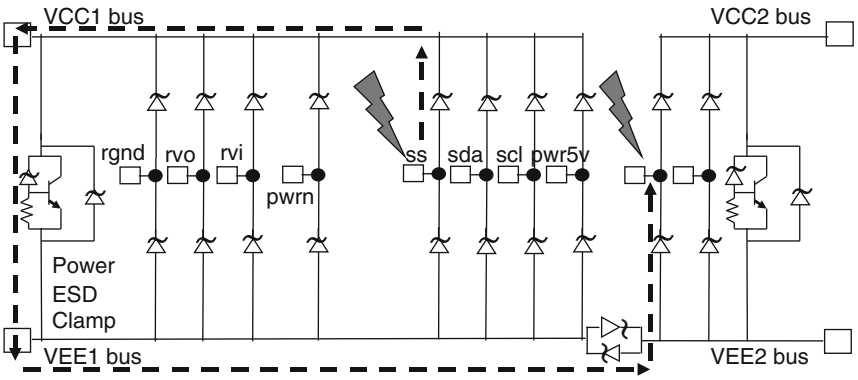


Fig. 6.10 Alternative ESD current paths in a two power domain amplifier network realized between different domain I/O pins

example, the total voltage for an ESD zap combination between the low and high-voltage inputs will accumulate the voltage drop on three diodes, in addition to the voltage drop on the busses and power clamps.

Thus, to provide a discharge current path for different input and output zap combinations, the diodes should be optimized both for low on-state resistance and low parasitic capacitance and leakage.

The total voltage drop at the output depends on the pin-to-pin combination (Fig. 6.10). Even for two voltage domains, the ESD current path may include current path scenarios with the clamp only, the clamp plus one diode voltage drop, +2 diodes, or +3 diodes.

If the ESD protection window is insufficient and separation resistors cannot be placed at the pins, then an additional local clamp, for example, one similar to the power clamp, can be applied at the critical pins.

In the case of high-voltage circuits, it is advantageous to use low-voltage ESD diodes in the back-to-back diode clamp between the domain grounds and at the differential inputs. The last requires an appropriate isolation of the diodes from the substrate in case of non-SOI processes. An additional benefit of this separate diode selection is the smaller footprint of the low-voltage back-to-back diode clamp (Fig. 6.10).

6.1.4 Bipolar Output Protection in Low-Voltage Amplifiers

An example of a low-voltage, high-speed amplifier with a BJT output and protection components is presented in Fig. 6.11a. This ESD protection solution has been implemented for the high-performance, current-feedback differential amplifier [105].

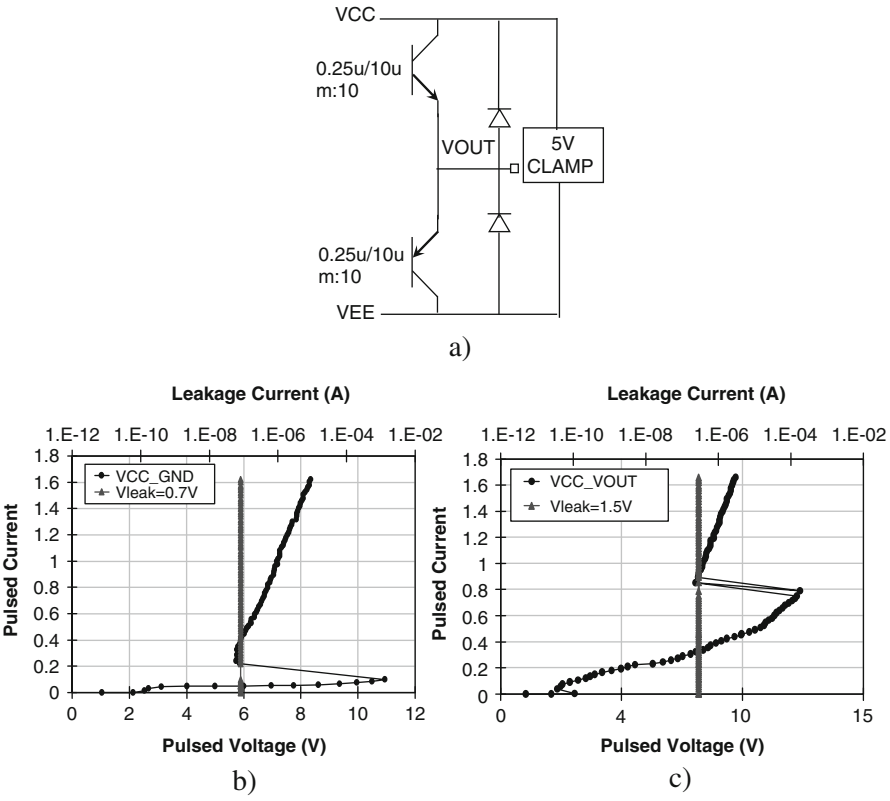


Fig. 6.11 An example of the BJT output in a high-speed amplifier (**a**); TLP characteristics for the power supply domain (**b**) and output to the ground (**c**) zap combinations

In this case, the total width of the 5 V output devices is $W = 200$ µm, which is sufficient to provide self-protection. With the particular design parameters, the selected FOX snapback power clamp has a rather high triggering voltage.

For the VCC to VEE zap combination, the clamp, with a triggering voltage of ∼11 V and a holding voltage of ∼6 V, conducts the ESD current (Fig. 6.11b).

However, in the case of the output VOUT to power supply VCC ESD stress combination, the TLP characteristics are different. The comparative analysis shows that in this case, the output NPN BJT transistor already provides an ESD current path to the ground while the power clamp remains passive. This conclusion about the current path can be easily confirmed by the observed different triggering and holding voltage levels (12.5 and 8 V, respectively) (Fig. 6.11c).

Thus, the circuit (Fig. 6.11a) relies on the self-protection of the BJT device of appropriate size for at least one current path.

This measure is often advantageous in practical amplifier designs. Even if NPN operation is unstable in high-current mode, its self-protection capability can often be improved by additional polyemitter ballasting resistors [115] that ensure the current balance in the BJT array (Fig. 6.12). A similar approach can be used for small devices to increase the ESD protection window. In this case, an extension of the pulsed SOA for small output BJT devices is used to ensure the turn-on of the ESD clamp.

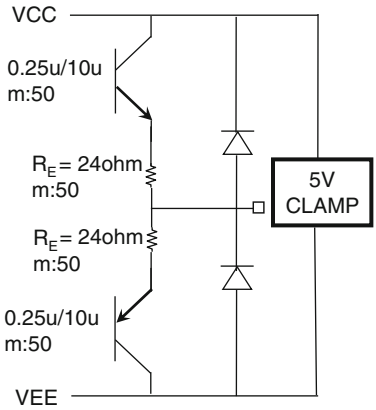


Fig. 6.12 Using additional polyemitter ballasting resistors in the 5 V bipolar output stage for the ultra-low-noise voltage-feedback operational amplifier

6.1.5 Input Protection

The BJT differential input is usually protected by the back-to-back stacked diode clamp (Fig. 6.13). To avoid pin mismatch, the clamp is composed of substrate-isolated p^+-Nwell and non-isolated n^+-Pwell diodes. The shallow junctions of the diodes are connected to the input pins in order to reduce the parasitic capacitance introduced by the diodes, while the well diffusion is connected to the floating nodes of the clamp.

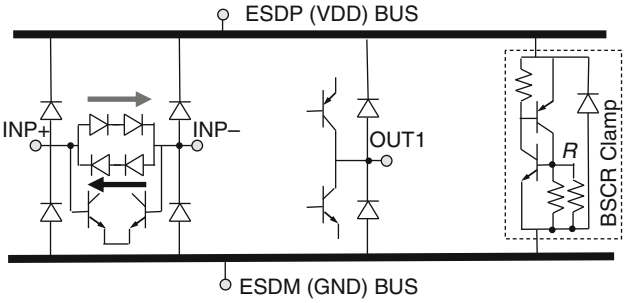


Fig. 6.13 Differential input protection

For the CMOS digital input, the most sensitive device components are the MOS devices, due to gate oxide overstress. In principle, rail-based protection can still be used successfully in this case, especially if the circuit's performance level can tolerate an additional decoupling resistor for the second stage (Chapter 5).

One of the most important aspects of successful ESD design is avoiding alternative damaging ESD current paths through the internal circuit.

This can be illustrated by an example of antenna diode placement. Usually, an additional minimum dimension region in the Pwell region of the NMOS device forms the antenna diode. The placement of the diode is a design choice. If this n^+-region is drawn too close to the source diffusion, a parasitic NPN with rather high base resistance may be formed (Fig. 6.14b). Depending on the process technology and the Pwell connection, if minimum process rules are applied, the triggering voltage of such an NPN device may become even smaller than the gate oxide breakdown voltage, thus providing a new severe limiting factor to the ESD protection window. To avoid such an issue, the antenna diode should be either spaced apart from the input NMOS source or an additional p^+ diffusion region should be placed closer to the antenna diode.

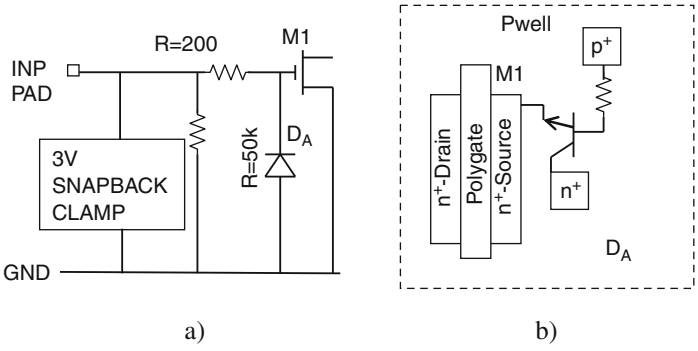


Fig. 6.14 Local protection of the CMOS input (**a**) and an example of the layout with an incorrect antenna diode placement (**b**) that enables a parasitic NPN BJT device with high base coupling

Similarly, an ESD protection window limiting can be caused by the MIM (metal–insulator–metal) capacitors used at the input and so should also be considered.

In the case of the low-noise amplifier (LNA) bipolar input, it may not be possible to implement a second-stage resistor due to design requirements. At the same time, minimally sized BJT devices may be used to provide the required circuit performance (Fig. 6.15). There are current paths in the circuit both from the input to the power supply (VCC) and from the input to the ground (VEE) that can be observed before the triggering of the local protection clamp, depending on the conditions of M3 and M4.

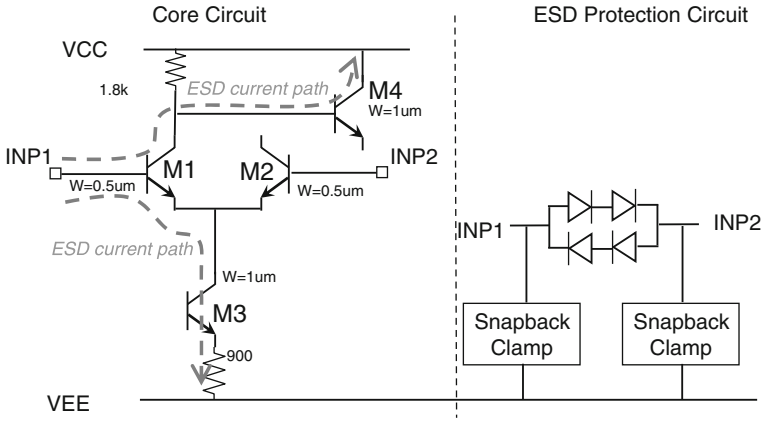


Fig. 6.15 An example of the minimally sized BJT LNA input circuitry with ESD protection and a possible additional current path scenario inside the internal circuit

A possible fix for this circuit protection issue consists of fine-tuning the triggering voltages of the ESD protection snapback clamps (Fig. 6.15), increasing the size of the input devices, and by additional resistor to collector of the BJT M3.

6.1.6 CMOS Output

Low-voltage CMOS output protection in amplifiers is similar to the strategies discussed in Chapter 5. If the design can tolerate the active clamp approach, then the voltage limiting of CMOS circuit are rather straightforward and similar to the protection of digital circuits.

As an option for CMOS output pins with an open drain, the NMOS can be redesigned with a drain ballasting region to match each drain finger, instead of the lumped separation resistor.

6.2 Digital-to-Analog and Analog-to-Digital Converters

The analog-to-digital converter (ADC) transforms (converts) an input analog voltage or current into a digital output proportional to the magnitude of the original voltage or current.

ADC's are widely used for processing an analog signal in a digital form. For example, ADC's are found in XM radios, digital video cards, microcontrollers, and digital oscilloscopes.

The serial digital data interface is used to reduce the cost of the products, by minimizing the pin count in ADCs.

Respectively, the functionality of the digital-to-analog converter (DAC) is the conversion of an arbitrary finite-precision number at a digital input into a corresponding analog output signal that represents a physical quantity, for example, voltage or current.

In the time domain, DACs are often used to convert finite-precision time series data into a continually varying physical signal.

A typical DAC conversion method is based on delta–sigma modulation for producing a pulse-density modulated signal that can then be filtered to produce a smoothly varying signal.

DAC applications cover a broad spectrum of systems: digital audio systems from players to sound cards; digital video systems and displays; and other consumer, industrial and medical applications.

To some extent, the basic ESD protection approach used for digital–analog converter (DAC) and analog–digital converter (ADC) products can be developed according to a general understanding of the corresponding functional blocks of the converters. At an even higher level, both DAC and ADC combine features that allow application of ESD design principles for digital blocks and the previously described analog blocks.

In most typical cases, the digital part can be protected using the rail-based active clamp approach with corresponding upper and lower ESD diodes from the I/O to the rails. In this case, the most attention is focused on meeting the CDM protection requirements and some of the ADC and DAC digital blocks.

The specific case of the serial data line pins (or I2C) has been already discussed above and are related to open drain protection.

There is a corresponding special case of ADC protection that is related to the low-voltage differential signal (LVDS).

However, to some extent, protection of the analog part in ADC and DAC can be treated somewhat similarly to the protection of the amplifier input and output, respectively. In a high-end, multibit ADC, the ESD/signal interference is the major challenge in providing an appropriate analog input signal resolution.

The analog output of the DAC can be treated as a CMOS or BJT output, similarly to the amplifier discussed in the previous section.

To illustrate the major principles of ESD protection for the DAC, the major functional blocks and most typical external pins of the 14-bit DAC are presented in this section.

The selected example is addressed for a DAC designed for the 0.13 μm DGO CMOS process with two voltage domains: the 1.2 V domain for the digital blocks and the 3 V voltage domain for the analog blocks.

6.2.1 Functional Blocks for High-Speed DAC

A simplified block diagram of the major functional blocks of digital-to-analog converter is presented in Fig. 6.16.

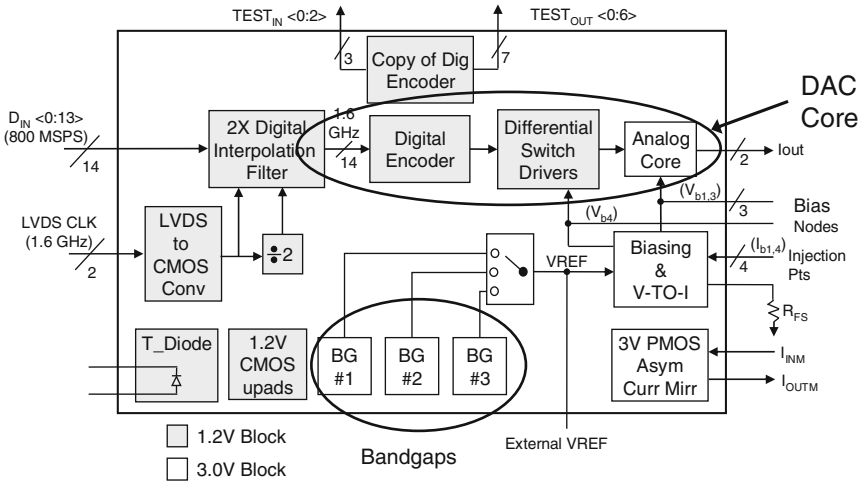


Fig. 6.16 Simplified block diagram for the digital-to-analog converter

The purpose of the system is to convert digital code received as serial low-voltage differential signal (LVDS) with 1.2 V amplitude into an output analog signal of up to 3 V (Fig. 6.16).

The 14-bid DAC contains 11 major functional circuit blocks. From the ESD protection point of view the most important information is about the devices and circuits directly interfacing with external pins. Practically all the digital parts of the converter are based on 1.2 V devices, while the analog parts are based on 3 V devices.

The functional circuit blocks are designed with different parts of the DGO CMOS process technology, using high-speed 1.2 V CMOS devices with thin gate oxides (NMOS and PMOS) and 3 V-tolerant thick gate oxide CMOS devices (HNMOS and HPMOS). The deep Nwell layer available in the process isolates the NMOS structures from the p-substrate, thus generating the corresponding isolated devices NMOSI and HNMOSI.

The biasing block is mainly based on the 3 V HPMOS and HNMOS. The block sets bias currents for the DAC and the LVDS-to-CMOS converter, based on an external bias resistor. This block requires the reference voltage V_{REF} to function.

The other 3 V pins are represented by the bandgap reference circuits based on HNMOS, HPMOS, lateral PNP devices, and Deep Nwell NPN devices. The PMOS current mirror is designed based on the HVPMOS and the PMOS with 3 V devices that interface with the external pins in the biasing circuitry block and analog core output.

The PMOS current mirror (Fig. 6.17) shows a single cell corresponding to DAC's output pins.

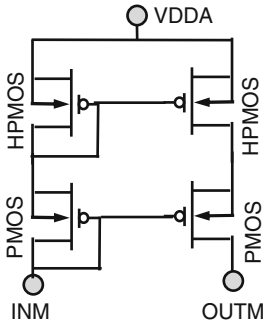


Fig. 6.17 PMOS current mirror

One of the three band gaps can be selected to produce a stable reference voltage V_{REF}. V_{REF} can be externally measured and used to compare the differences in the three band gaps. The band gap block requires the mode control block to function. The block is built on the 3 V devices (Table 6.3) with the external pins VREF, analog VDD (VDDA), and analog ground (AGND).

Table 6.3 Waveform parameters for short-circuit load conditions

Stress level	Simulator voltage (kV)	Cable length (m)	Ipeak ±10% (A)	Rise time (nS)	Pulse width ± 10%(μS)
2	+1.0	100	4.5	1–10	0.8
4	+2.0	100	9	1–10	0.8
2	+1.0	200	4.5	1–10	1.6
4	+2.0	200	9	1–10	1.6

The digital encoder converts a 3-bit binary code into a 7-bit code. The code used by the DAC for MSB and LSB segments. The digital encoder relies on the *LVDS-to-CMOS converter* for the clock signal. The corresponding external pins are Test_in and Test_out, digital VDD (VDDD), and digital VSS (VSSD).

The mode control block selects one of the three band gap references to use as a reference for the DAC. This reference also provides power down conditions for the DAC. The mode control outputs are internal signals only and thus do not require separate ESD protection, as they rely on the ESD protection already provided to the band gaps and the DAC. The 1.2 V NMOS and PMOS devices are used in the block.

The 1.2 V LVDS-to-CMOS converter converts a low-voltage differential input signal (LVDS) into a single-ended output signal. Only internal circuit uses this output signal. The block requires mode control and biasing blocks to function.

The 1.2 V digital interpolation filter circuit block doubles the input data rate by linear interpolation. The outputs of this block are only connected to the DAC. The interpolator requires the LVDS-to-CMOS converter to function.

The DAC core produces a programmable current that depends on the input code of the DAC. This input code is generated by the interpolator. The core requires external 50 Ohm resistors to the AGND on the IOUTP and IOUTN.

6.2.1.1 ESD Protection Network

A two-power domain ESD protection network is implemented with two digital (DGND) and analog (AGND) decoupled ground busses. One of the most effective ways to protect the DAC with two-voltage domain is the rail-based active clamp ESD design.

In most practical cases, the high pin count of the chip and the multiple (over 7) layers of thick bond metal or copper top metallization provide appropriate process capabilities for the rail-based active clamp network.

Practically, the local snapback ESD protection is required for the DAC only in the case of high voltage, for example, in a charge pump or for EEPROM pins (as well as certain specified system-level pins).

In principle, the ESD protection network for the DAC can be composed from either the two-domain active clamp clusters or use only HV devices provided by DGO process.

In the case of the double voltage domain approach, the 1.2 V digital and 3 V analog circuit blocks require two separate rail-based ESD network domains based on the 1.2 V active clamp and the 3 V active clamp.

For high-performance, the low-voltage domain should be designed as a substrate-isolated circuit. This is achieved using DeepNwell isolation. A corresponding specific for active clamp design is also related to substrate isolation.

An example of the principal circuit design for a low-voltage active clamp in the case of a 0.13 μm LV CMOS device is based on a 1.2 V isolated NMOS with W/L_G = 500 μm/0.13 μm (Fig. 6.18). This design can provide an on-state resistance as low as R_{ON}=4 Ω and a leakage current of I_{DSS}=1 μA at the nominal VDD power supply voltage level $V_{DS} \sim 1.2$ V.

This high leakage current is present due to the utilization of 1.2 V devices and may be considered too high in some cases. The alternative is to use the 3 V active clamp with a much lower leakage.

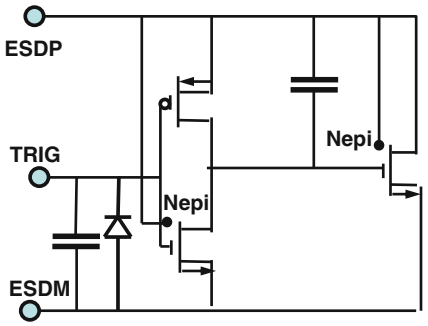


Fig. 6.18 Schematic for isolated 1.2 V slave active clamp

The high-voltage 3 V active clamp design is similar to the already discussed solution in Chapters 4 and 5. In the case of the 3 V CMOS components with W/L_G=500 μm/0.3 μm, the HVNMOS device can provide a on-state resistance below 10 Ω and a leakage current below 50 nA at a nominal VDD power supply level V_{DS}=3 V.

Based upon these parameters, an ESD protection network with only one type of the 3 V active clamp is sufficient if there is an adequate number of pins. Such a design provides a significant simplification of the chip layout, especially if the sequence for the pins for HV and LV domains is mixed in the pad ring.

In this case, the power supply and the other low-voltage domain pins are simply treated similarly to the I/O pins in the 3 V domain, with a corresponding connection of the upper and lower ESD diodes to the 3 V ESD plus (ESDP) and ESD minus (ESDM) metal busses. This design imposes a less stringent requirement for the low-voltage power supply and ground busses and relies on the rather low-voltage clamping level provided by the HV Active clamps.

In addition to the "slave" active clamps connected to each pad, four RC-timer corner cells are embedded in the circuit to complete the distributed ESD network and control the trigger line potential. The RC-timer cell provides the trigger-off signal with a delay time of ∼10 μs.

For successful implementation of the design, one of the most critical design items is the ESD bus. Ideally, the ESD busses use all layers of metal, which reduces the accumulated resistance across the whole chip, since the bond pads can only use three or four upper metal layers.

The full test chip simulation result for the 140-pin corner-to-corner ESD bus resistance is below 1 Ω. At a 3 kV HBM ESD pulse, the device gives only a 1.2 V peak voltage drop across the ESDP and ESDM buses.

The CMOS input protection circuit is similar to all I/O circuits. It combines a two-stage network with a 100-Ω polyresistor to provide sufficient CDM immunity. Each input gate of the circuit is tied to the local VDD and VSS circuit nodes with minimally sized additional diodes (Fig. 6.19).

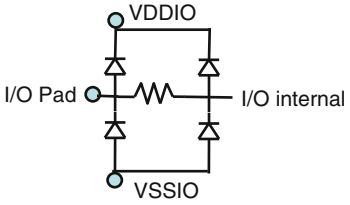


Fig. 6.19 ESD protection for I/O pin with gates tied directly to the pin

6.3 High-Speed Interface IO pins

6.3.1 Interface Analog Products

The main function of the interface circuit is to provide a datalink between different systems or subsystems. There are a number of interface standards that cover different datalink specifications for transmitting signals through a cable, from several inches within the same system block to several thousand meters between systems within the same building or facility.

In comparison with DAC, one the major specifics of advanced interface products is the high speed of the data transmission (up to 10 Gbit/s and above).

From an ESD specific point of view, a typical requirement may include on-chip system-level pins. Typically, the cable discharge event (CDE) (Chapter 8) protection level is specified. Another peculiarity of the interface circuit is the presence of high common-mode voltages in the input signal.

Just like RF-amplifiers, the high-speed input interface pins require broadband characteristics [106]. This eliminates the possibility of application of a number of narrow band, LNA-specific input solutions that rely on ESD pulse bandwidth filtering [116].

Over recent years, interface applications are heading toward 8–12 Gbps links. Achieving these goals requires not only increase of speed but also corresponding reduction in power consumption and noise.

Typically, interface parameters are related to the power moving in the network and in and out of the integrated circuits. The interface characteristics are often described in terms of the equivalent S-parameters.

The *transducer gain* parameter G_T represents the ratio of the power delivered to the load to the power available from the source.

At matching load and source impedance, $Z_L = Z_S = 50\ \Omega$ and $G_T = |S21|$.

The *operating power gain* parameter G_P represents the ratio of the power delivered to the load to the power delivered to the network.

The power delivered to the network is dependent on the S11 parameter. In this case, $G_P > G_T$ and G_T is closer to G_P with better input matching.

The *available power gain* parameter G_A represents the ratio of the power available from network to the power available from source.

Since the power available from the network is dependent on S22, $G_A > G_T$, while G_T approaches G_A with better load matching.

A functional block diagram an interface application is presented in Fig. 6.20.

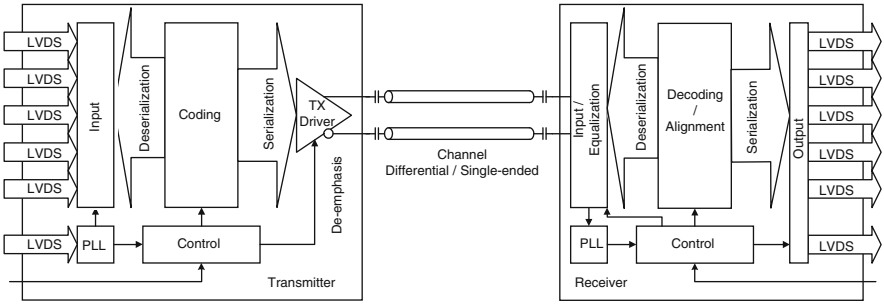


Fig. 6.20 Function block diagram, for example, of LVDS interface solution

In terms of the ESD protection challenge, the interface application requires the CDE protection for the selected pins directly interfacing with cable. To address the design specific, the next section presents the CDE test methodology.

6.3.2 Cable Discharge Event Test Procedure for Integrated Circuits

The cable discharge event (CDE) is a form of system-level stress. System-level protection requirements are discussed in Chapter 8. CDE is very specific to interface products.

CDE can be a significant threat to Ethernet transceivers because of the long cables involved. Until recently, there was no industry standards and limited literature available on CDE [117–120]. Measured CDE waveforms are obtained by charging short USB cables and manually discharging them. The discharge waveforms showed a fast initial overshoot followed by a rectangular pulse. The rise time of the pulses was estimated to be a few hundred picoseconds. In practical cases, due to the inherent properties of air, discharge repeatability cannot be expected in manual cable plugging.

The test procedure for integrated circuits was originally developed by industry leaders [121].

The CDE testing is categorized as either *Common-Mode Charging* (CMC) or *Differential Mode Charging* (DMC). At CMC, all the wires in the cable are charged simultaneously with respect to a reference. At DMC, the individual wires in the cable are charged with respect to each other.

At pin sequencing, some of the pins of the cable are connected to the circuit prior to the other pins, thus resulting in a differential mode signal at the Ethernet transceiver.

A typical waveform for CDE is presented in Fig. 6.21. Rise time is defined as the time required for the leading edge of the pulse to increase from 10 to 90% of its

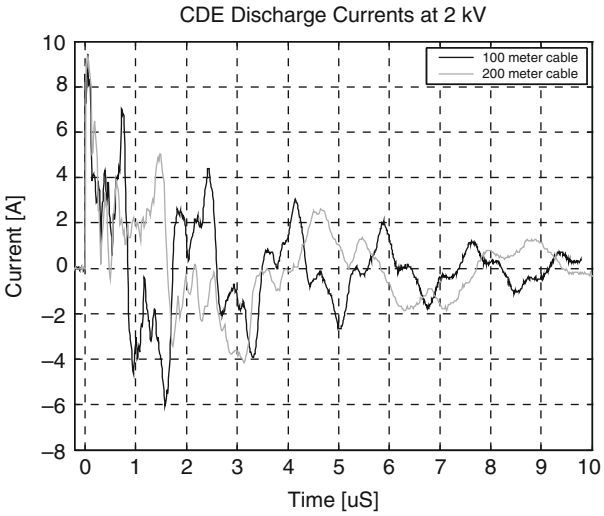


Fig. 6.21 Typical discharge current waveforms measured for a short-circuit load [121]

maximum value. Pulse width is the time interval between the start and end points of the first peak of the discharge current waveform. The peak current I_{peak} in Fig. 6.21 is the maximum current.

Typically, the ESD simulator is equipped with a 200-m long CAT5 Ethernet cable, a 100-m long CAT5 Ethernet cable, a high-voltage power supply, and a relay box. The testing may be done with either one or both cables depending on the test requirements.

The collected discharge current waveforms should comply with the current waveforms specified in Fig. 6.21 and by Table 6.3 CDE testing is conducted for voltage levels with a 500 V step, with Level 1 of +/– 500 V up to Level 4 of +/– 2 kV.

The test setup is shown in Fig. 6.22. In this example, the CDE stress pulses are applied to the DUT (device under test) soldered onto a test printed circuit board (PCB). The DUT can also be mounted into a socket on the test board.

The device pins receiving CDE stress (four pins or eight pins depending on the Ethernet standard used) are accessed through an RJ-45 connector on the board. To simulate the worst-case stress on the IC, the IC pins should be directly connected to the RJ-45 connector without any magnetic or discrete components.

First, the charging relay is momentarily closed and opened to charge the Ethernet cable. Then, the discharge relays in the relay bank are closed all at once or randomly, depending on the test requirements.

Depending on the Ethernet standard used by the DUT, four of the discharge relays may not be switched. For example, in the case of a 10 Base T Ethernet standard, only four pins of the DUT are stressed by the closing of four relays.

Prior to testing the device, waveform verification should be performed at 2 kV level with the 200-m long cable. A minimum of three samples should be tested

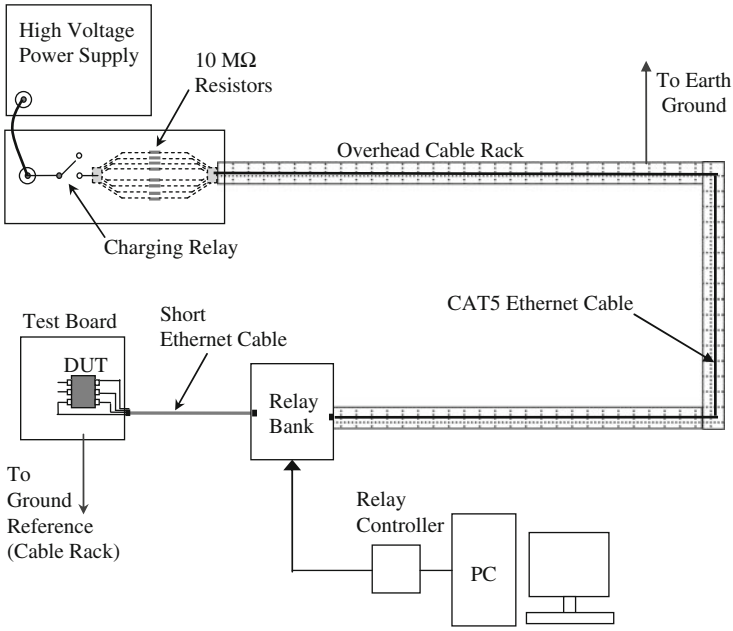


Fig. 6.22 CDE simulation test setup [121]

at each required voltage level. Ten positive and 10 negative discharges should be applied to all Ethernet ports of the device for each voltage level. A ground terminal on the test board must be connected to the ground reference (the cable rack).

As a system-level test, CDE testing should be conducted in both the powered and the unpowered states of the DUT. In the powered mode of testing, the vulnerability of the DUT to the latch-up phenomenon should be monitored. For that purpose, the DUT should be powered at the maximum nominal power supply level using external power supplies. If multiple supplies exist, all supplies need to be turned on. The supply current(s) must be monitored after each stress to determine whether the DUT is latched up. If the latch-up phenomenon is observed, the device is considered to have failed at that stress level.

Each device must pass the electrical, parametric, and functional tests specified in the device data sheet after exposure to CDE pulses. Each device also must not show any latch-up during exposure to CDE pulses.

In case the test setup for addressing the system-level on-chip requirements described above is lacking, elevated package-level requirements are often specified. For example, instead of a standard package-level HBM 2 kV, an 8 —15 kV HBM can be specified for the selected interface circuit pins.

Alternatively, cable interface pins can also be tested for meeting the IEC 61000-4-2 8 kV contact and 15 kV air-gap specifications. However, the charge cable test is still the main verification tool.

6.3.3 ESD Protection of Interface Pins with CDE Requirements

Typically, for multiple-pin-count interface products, an elevated level of ESD protection can still be achieved with the active clamp rail-based approach by a simple scaling of the number of clamps.

General guidelines for implementing system-level protection include connecting the "heavy-metal"/ESD diodes to the external decoupling caps. To achieve local clamp, high-current-level SCR protection for power-on ESD, the SCR holding voltage must be higher than the power supply voltage.

This alternative design includes high holding voltage dual-direction clamps described in detail in Chapters 3 and 4. Wide common-mode range standards require application of the isolated NMOS and LVTSCR ESD clamps, as well as the dual-direction SCR's (DIAC) structures.

The dual-direction voltage tolerance is determined by the signal standards. For multipoint LVDS (MLVDS), the common-mode voltage range is $V_{CM} = -1.4$ to +3.8 V, while the signal voltage $V_{\text{signal}} = -2.6$ to +5 V on a 3.3 V supply. For the eXtended LVDS (XLVDS) interface standard, $V_{\text{CM}} = -5$ to +12 V and $V_{\text{signal}} = -6.2$ to +13.2 V on a 3.3 V supply. Other typical standards are the high-speed HS (2.5 V) and a slow SMBUS.

Often, for a typical ESD library designed for compact mixed-signal applications in the absence of the CDE tester, a correlation between the package-level test pulses can be assumed followed by the agreement with the customer upon the alternative test specifications. For example, an ESD library compact mixed-signal application can be designed to meet >8 kV HBM, 300 V MM, and 1250 V CDM.

The most challenging problem for high-speed applications is the low capacitance of <250 fF of analog IO pads. The pads are protected by ESD diodes that utilize a micro-SMD bump metal or inductor metal to achieve low capacitance and low resistance.

The cable interface pins are designed to withstand an ESD discharge with power-on. Due to the high CDE current, LVTSCR devices with holding voltages $V_{hold} > V_{supply}$ are used. The target performance is verified by the IEC61000-4-2 8 kV contact/15 kV air-gap tests. Similarly, solid metal/diode paths to external decoupling caps are required, while the low parasitic capacitance requirement can be below 600 fF.

At the same time, in opposite to the interface pins, the power supply does not require system-level protection. Active clamp protection on power supply pins and non-cable interface pins are tested at the package level in power off conditions.

Thus, to protect against CDE, specialized cells are needed to satisfy ESD protection requirements for both power-off and power-on conditions.

Thus, interface pin protection requires rather challenging device-level ESD solutions. Such solutions compose the system-level capability with a high holding voltage, thus incorporating an anti-latch-up feature on the device level; for MLVDS and XLVDS, the solutions should be dual-direction and provide low parasitic capacitances. At the same time, the clamp should be capable of withstanding high-speed

CDM. If the process supports only a 2.5 V device, the 3.3 V I/O and ESD protection can be based on stacked ESD devices.

In practical design work, such solutions are usually achieved through extensive experimentation cycles. Often, the solution demands additional space on the chip and may be process-sensitive.

The size of the ESD diodes in high-speed I/O can be reduced by 40%, thus enabling aggressive interface product design.

For the BiCMOS process, the use of the integrated back-to-back dual-direction devices of superior performance and a small footprint is described in Chapter 4.

Often CDE level can be achieved using conventional active clamp protection network. An example of the CDE protection for Serializer and Deserializer over a single differential pair is presented in Fig. 6.23. The product example is specified to work with cable up to 10 m and designed for automotive video displays with LVDS standard compliant to ISO10605 "Test methods for road vehicles – electrical disturbances from electrostatic discharge."

In addition to standard package level ESD protection requirements system-level ESD protection is built into the IC. High cable discharge level protection is verified by ISO gun +/– 10 kV contact and +/– 30 kV air discharge between "pin under test" and ground. The system-level pins are the LVDS input and output of the single differential pair.

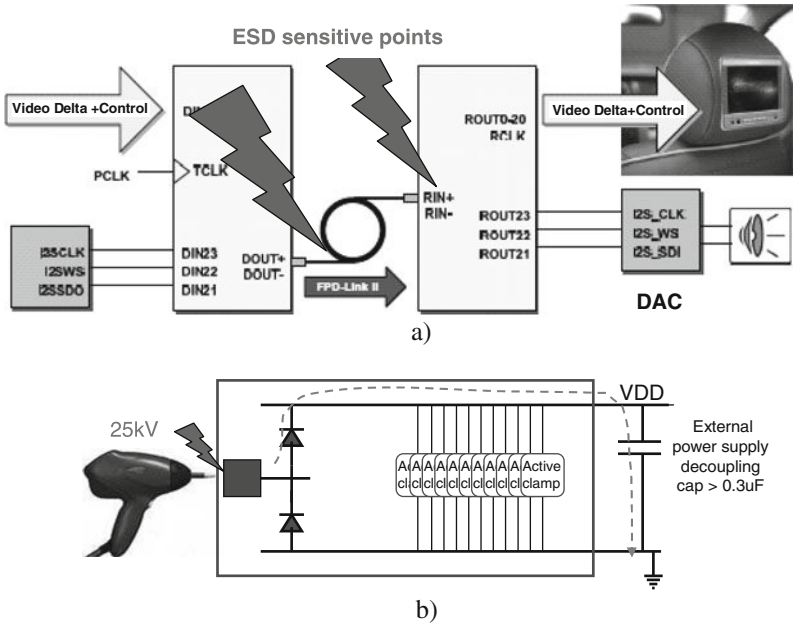


Fig. 6.23 Example of interface application network (**a**) and ESD protection scheme with application board bypass capacitor (**b**)

An important part of the protection system and the testing setup itself is application board capacitors. In this particular example 0.5 μF bypass capacitor was part of the system and thus it has been included for ESD evaluation too (Fig. 6.23b). At the same time the ESD scheme rely on distributed active clamp approach with specially designed ESD diodes capable of handling very high pulsed currents of ~30 A.

6.4 Summary

In this chapter, ESD protection challenges and solutions have been demonstrated for signal path analog circuits. In the case of signal path products, the ESD protection network is based on the conventional rail-based or local clamp approaches; the specifics of ESD protection are related to the product-specific pins.

When the active clamp rail-based approach cannot satisfy the requirements, major help is expected from device-level ESD solutions for the dedicated pins. The main challenges of ESD protection are in the minimization of the parasitic influence of the ESD components on the input signal.

The high pin count signal path analog product packages create challenges for the CDM ESD pulse protection. Since in high-speed circuits the second-stage resistor cannot always be of a desired value, the design effort is focused on the appropriate layout and schematic design of the I/O itself.

While the DAC and ADC can be protected in most cases by applying the same principles as described for amplifiers, the interface circuits bring in the system-level CDE protection requirements for pins interfacing with external cables and transformers. In high data rate pins of 10 Gb/s and beyond, the device-level ESD protection solutions are perhaps the most important methodology to understand and implement.

The device-level positive and negative feedback for implementing such solutions is discussed in Chapters 3 and 4. The practical design work of such complex solutions is usually achieved through extensive experimentation cycles. To reduce the time-to-market for such R&D work, physical device and mixed-mode simulations are critically important tools. Some of the simulation examples related to this chapter are presented below.

DECIMM^{TM} Simulation Examples for Chapter 6

To download a trial version of the numerical simulation software and request an electronic license key please visit http://www.analogesd.com

To download libraries with simulation examples for this chapter please visit http://www.analogesd.com/Chapter6.html

List of examples is subject to change.

Example 6.1a–6.1c Rail-Based Protection with Active 5 V NMOS Clamp and ESD Diodes

Library Name: Examples6_Signal_Path_Analog_Circuits
Project Name: E6.1a_Rail_Protection_5 V_AClamp_IOIO
Project Name: E6.1b_Rail_Protection_5 V_AClamp_ESDPIO
Project Name: E6.1c_Rail_Protection_5 V_AClamp_IOESDM

This set of examples (Fig. E6.1a) illustrates the current path and waveforms observed at different pin-to-pin combinations in a simple rail-based ESD protection network with two IO pins and an active NMOS clamp. The HBM waveforms (Fig. E6.1b) simulated for the example circuits can be compared to results obtained in the following example for the rail-based network protected by a 5 V snapback NMOS clamp. A separate analysis can be done for different designs of the Nwell and Pwell ESD diodes.

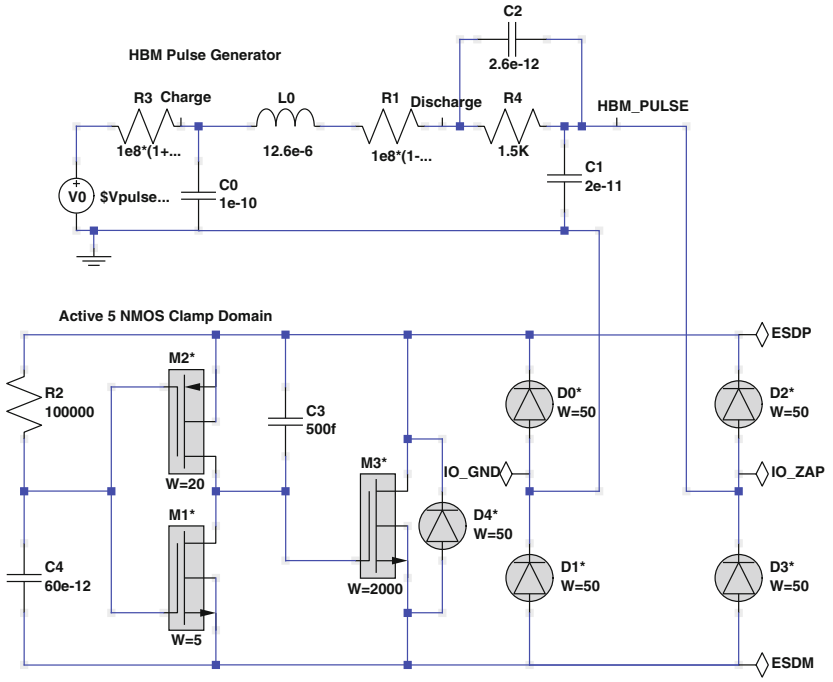


Fig. E6.1a Mixed-mode simulation circuits for different pin-to-pin combinations (IO–IO, IO–ESP, and IO–ESDM, respectively) and cross-sections of the finite element devices (5 V PMOS, 5 V NMOS, and Nwell and Pwell diodes, respectively) used in the circuits

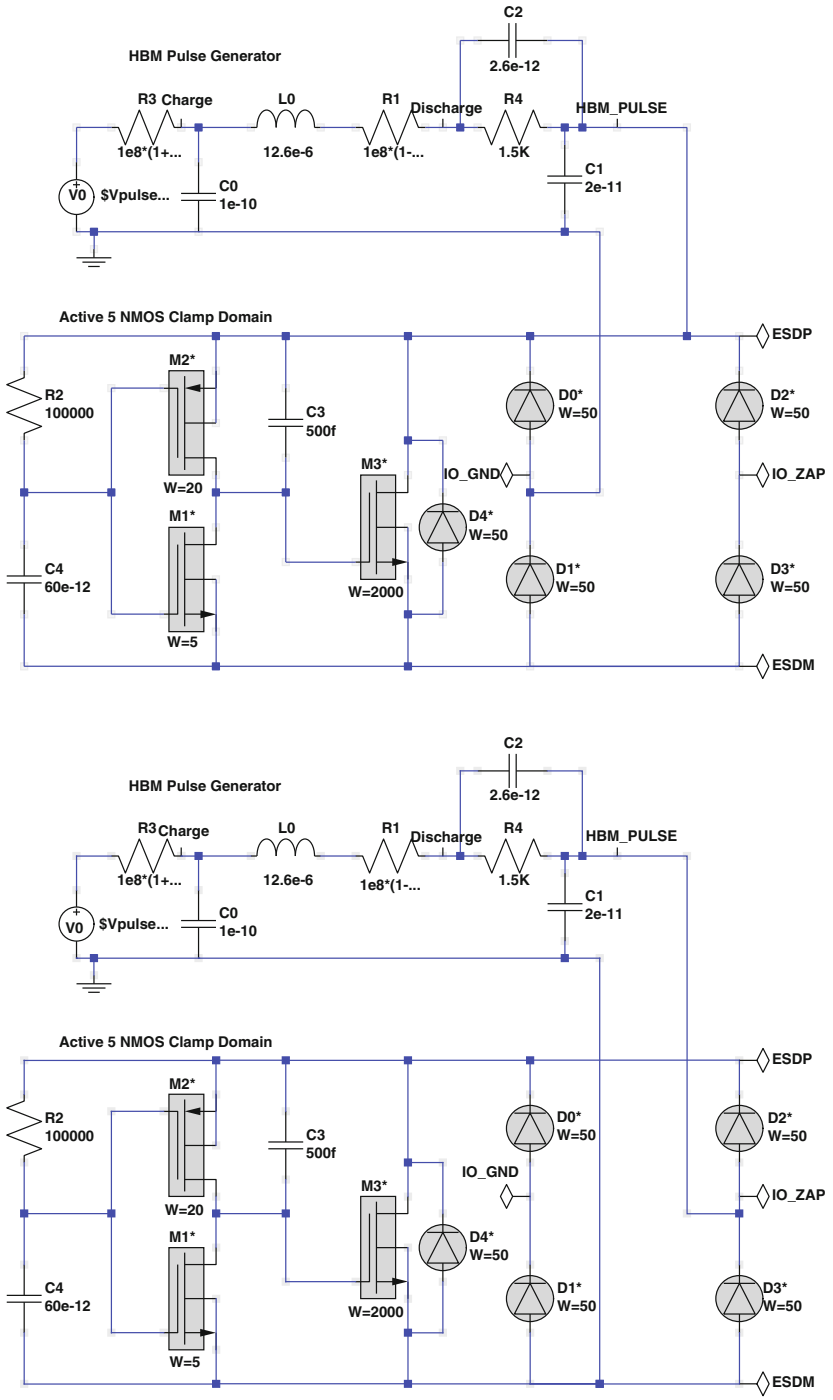


Fig. E6.1a (continued)

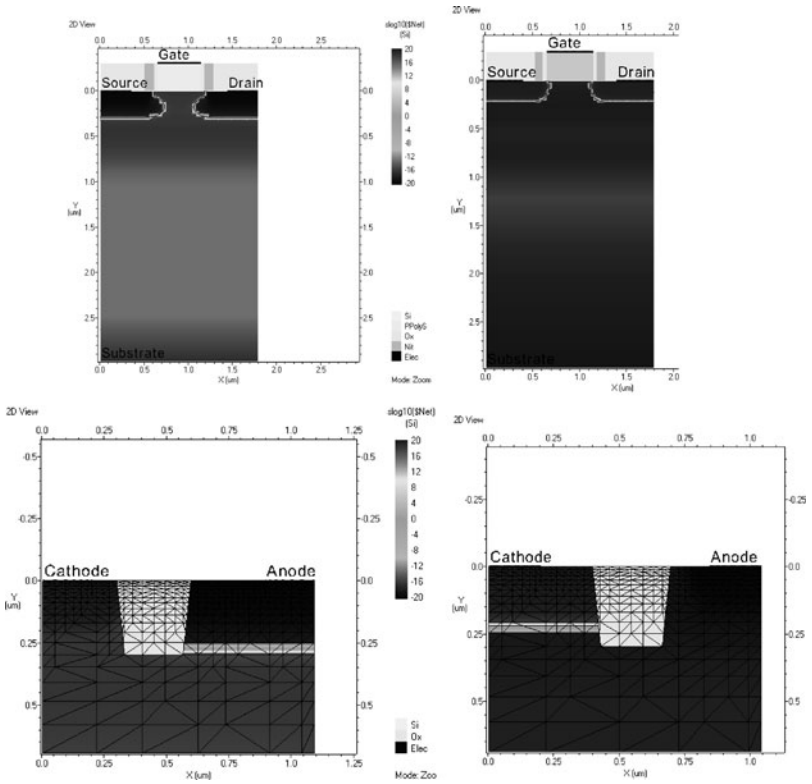


Fig. E6.1a (continued)

Example 6.2 Rail-Based Protection with 5 V Snapback NMOS Clamp and ESD Diodes

Library Name: Examples6_Signal_Path_Analog_Circuits
Project Name: E6.2_Rail_Protection_5 V_Snap_IOIO

This example (Fig. E6.2) illustrates the current path and waveforms at different IO and power pin combinations in the case of rail-based protection with an active clamp. The HBM waveforms simulated in this example can be compared to the corresponding waveforms in the above example for the rail-based network protected by the 5 V active NMOS clamp. This example can be used as a basis for analysis of different designs of the Nwell and Pwell ESD diodes and different snapback clamps, for example FOX and LVTSCR.

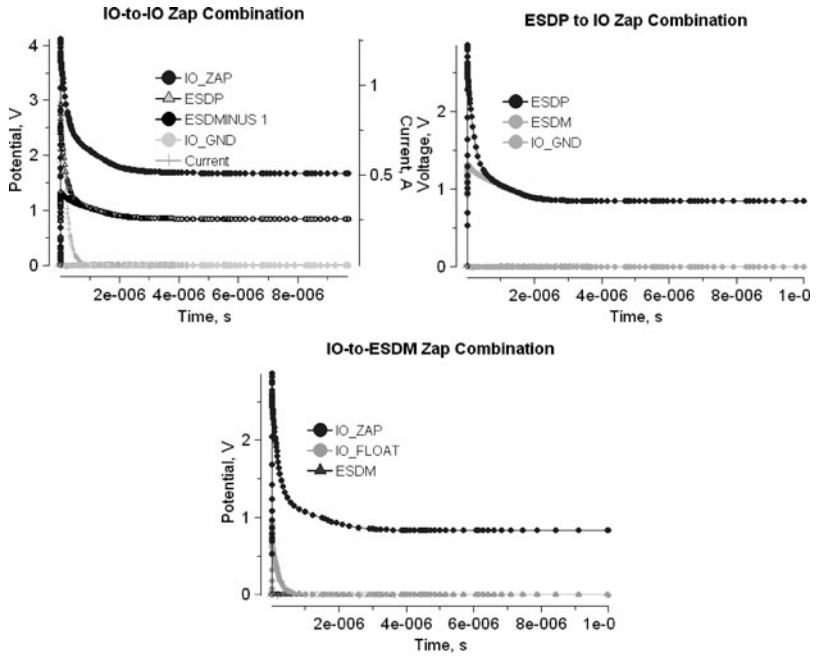


Fig. E6.1b Voltage waveforms at different circuit nodes for different zap combinations

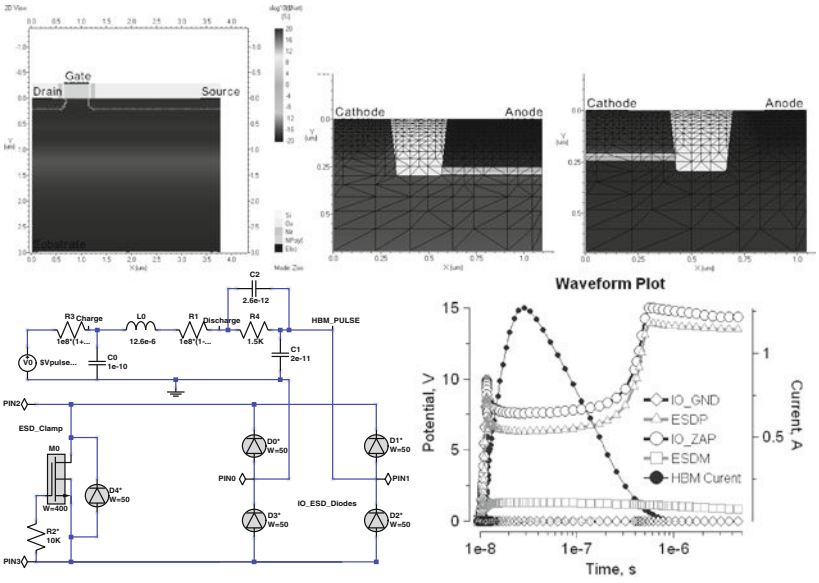


Fig. E6.2 Cross-sections of the finite element devices (5 V SNMOS, and N-well and P-well diodes, respectively), mixed-mode simulation circuit with two IO pins and snapback NMOS device and voltage waveforms at different circuit nodes under 2 kV HBM pulse

Example 6.3 Trans Impedance Amplifier

Library Name: Examples6_Signal_Path_Analog_Circuits
Project Name: E6.3_Trans_Impedane_Amplifier

This example (Fig. E6.3) is assembled using 1.5 V compact model (BSIM3) devices and demonstrates amplification of the low input current. An ESD protection network can be added to the circuit to study the voltage overstress at different amplifier circuit nodes. Amplification of 500 nA input current in the nominal operation regime is demonstrated in Fig. E6.3.

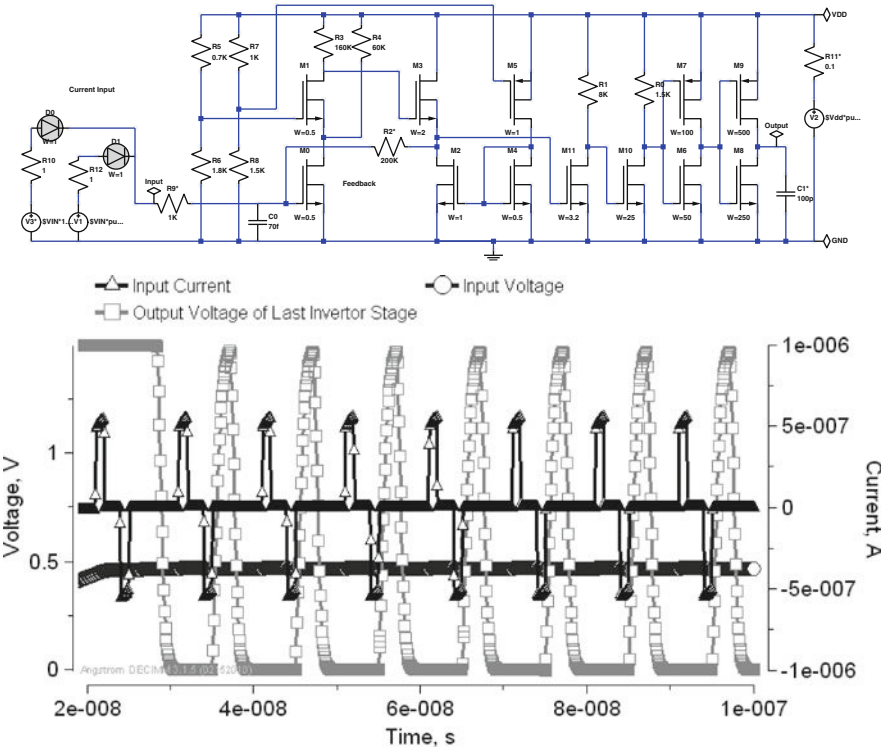


Fig. E6.3 Mixed-mode simulation circuit for compact model trans impedance amplifier and voltage and current waveforms at different circuit nodes

Example 6.4 CMOS Output Stage ESD Case

Library Name: Examples6_Signal_Path_Analog_Circuits
Project Name: E6.4_CMOS_Output_Stage_Case

This example (Fig. E6.4) presents a typical case of the CMOS output stage. The ESD current path can be studied as a function of the drain ballasting resistor values R2 and R5. The ESD current through the output stage usually results in irreversible

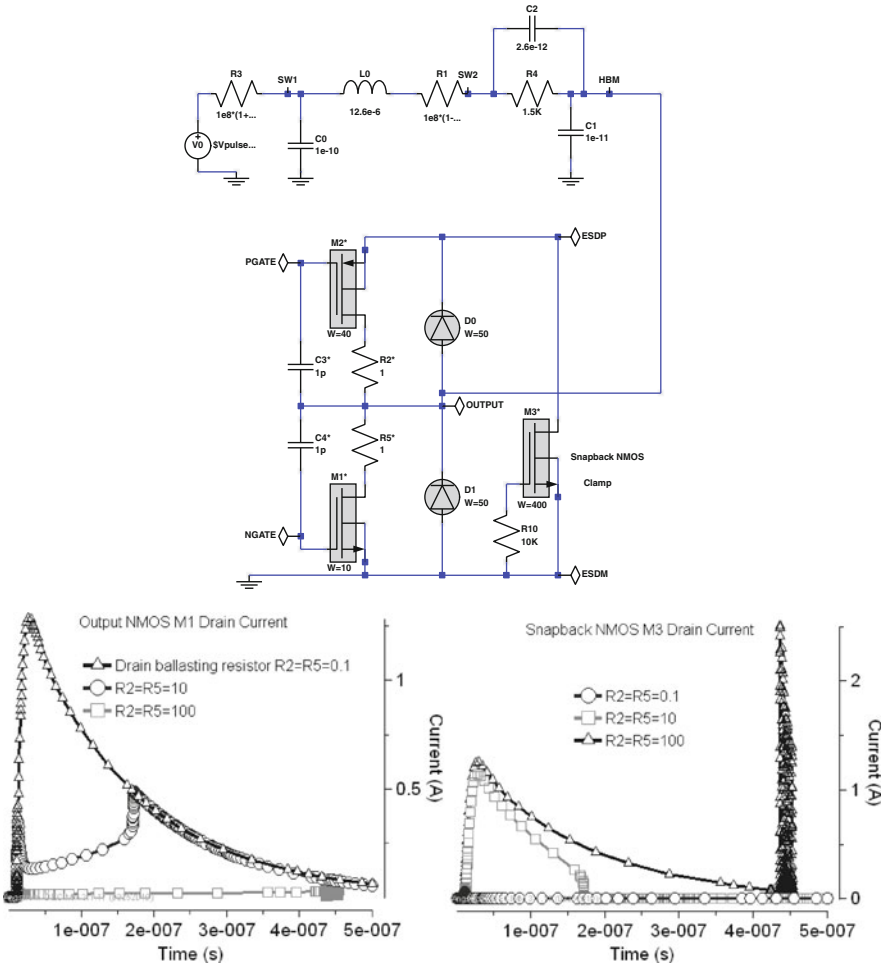


Fig. E6.4 Mixed-mode simulation circuit for CMOS output case and the waveforms illustrating the effect of the drain ballasting resistors on the current path

burnout of the output devices. The circuit can be used to study different pin-to-pin combinations and the output PMOS breakdown case in particular.

Example 6.5 CMOS Open Drain Case

Library Name: Examples6_Signal_Path_Analog_Circuits
Project Name: E6.5_Open_Drain_Case

This simulation example (Fig. E6.5) illustrates a current path alternative to the clamp (M2) through the internal circuit device (M1) as a function of the values of circuit elements C4, R5, and clamp resistor R10. Typically, the current path through the internal NMOS device indicates irreversible failure and the current through the

clamp should guarantee reversible ESD operation. At the same time, application of the snapback NMOS as M1 may provide self-protection of the entire circuit. Both scenarios can be explored in this simulation example.

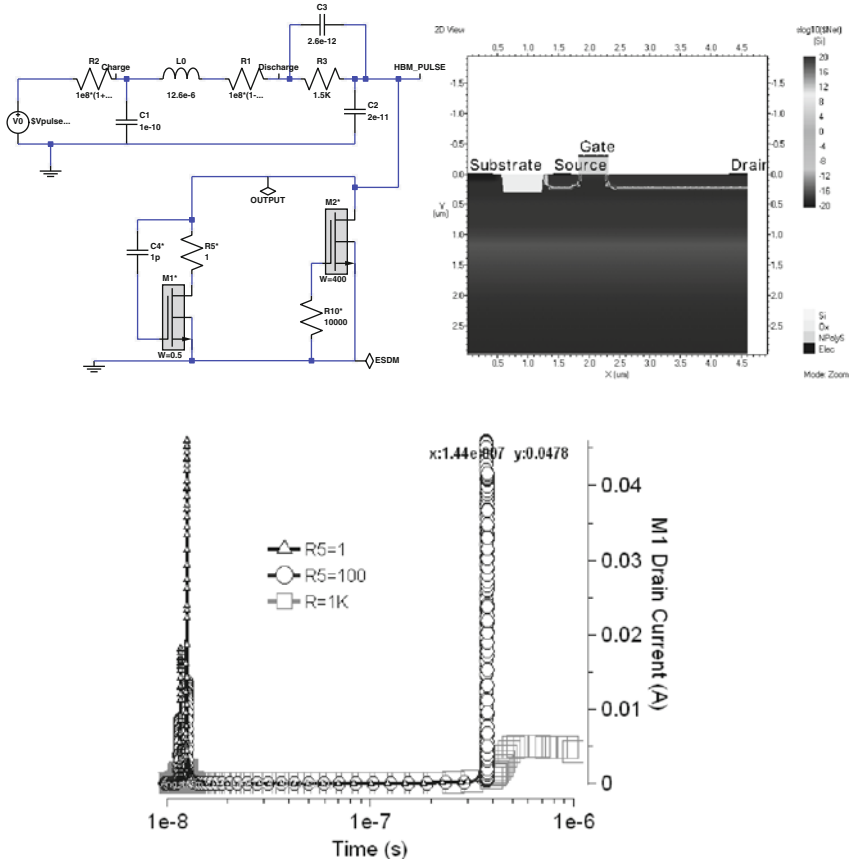


Fig. E6.5 Mixed-mode simulation circuit, snapback NMOS cross-section and current waveforms through the open drain device M1 under 2 kV HBM pulse

Example 6.6 BJT Output Stage Case

Library Name: Examples6_Signal_Path_Analog_Circuits
Project Name: E6.6_BJT_Output_Stage_Case

This simulation example (Fig. E6.6) for the open drain case demonstrates the current path through the internal circuit device Q1 as a function of the emitter ballasting resistor and clamp Q0 parameters (Fig. E6.6). This case covers output to VCC zap combination. The typical practical scenarios may include both irreversible failure and self-protection operation of the Q1.

Chapter 7
Power Management Circuits' ESD Protection

This chapter discusses the applications of ESD network and clamp principles, using examples of different power management analog circuits. The trends in the application of integrated power products are discussed in Section 7.1. Then, a more detailed analysis of both the most common and the most "aggressive" ESD protection solutions is presented in the subsequent sections with examples of cases concerning integrated power products, controllers, and LED drivers.

The specifics of power circuits significantly impact the chosen ESD protection strategy. Fast-switching and high-voltage pins, with their very limited ESD protection windows, create a significant challenge and often require a custom, circuit-dependent approach. Self-protection aspects of integrated power devices in complex, multiple power domain networks are hard to simulate. Substantial product-level experimentation is usually required to match power-optimized output devices with the proper protection solutions. These devices usually provide a rather narrow and sometimes even negative ESD protection window (in its classical understanding).

Similar to signal path analog circuits, the major principles of local and rail-based ESD protection network design, as well as ESD clamps architecture, apply to power analog products. However, a new specific is generated by the multiple current path scenarios, which can be realized in the power-optimized analog circuit due to the interaction of the internal components with the power pins.

In the case of fast transient pins connected to output devices with relatively low absolute maximum voltages, a significant challenge is related to the design and selection of the local clamps. In this case, a precise triggering voltage reference is the most desirable feature, which helps to minimize the dV/dt triggering effect.

The possibility of transient latch-up on switching pins or with a hot plug-in specification also requires countermeasures. Among them are the design and application of rather challenging high holding voltage solutions and the utilization of the self-protection capabilities of the internal circuits. These measures are especially important in high-voltage circuit cases. At the same time, realization of the two-stage ESD protection network becomes rather challenging due to performance losses with the additional resistance in the current path.

These and other aspects of power circuit ESD protection are addressed in the following sections by means of a more detailed analysis and case studies of the

V.A. Vashchenko, A. Shibkov, *ESD Design for Analog Circuits*,
DOI 10.1007/978-1-4419-6565-3_7, © Springer Science+Business Media, LLC 2010

selected specific analog circuits. This is done in order to create a substantial basis of general understanding that can be subsequently applied to arbitrary power analog circuits.

7.1 Power Management Products

7.1.1 Power Management Products and ESD Challenges

7.1.1.1 Market Trends

Understanding of ESD design for power products cannot be complete without consideration of the process technology specifics. One of the major trends for power management circuits today is their integration into practically all of the analog circuits.

DC/DC converters are usually a part of the power modules used today in telecommunication and data communication, computer and office automation, and industrial control and instrumentation. The fastest growing areas of use today is automotive electronics and general automation.

Voltage regulators are used in consumer electronics, computer and office automation, and tele/data communication.

Pulse width modulation (PWM) and PFM controllers are used in merchant DC/DC converters, computer and office automation, consumer electronics, general automation, and industrial control and instrumentation.

One of the major trends in power analog products is the increase in the output current of the integrated power regulator.

A current level of 4–8 A is used for point of load (POL) regulation; notebook computers; portable instruments, distributed power systems, battery powered equipment; high-end printers; DSP, memory supplies; automotive; distributed 2.5, 3.3, 5 V power systems; cellular base stations; xDSL modems. Even higher current (10 A and above) is required for ASIC, DDR memory; fiberoptics power supplies; telecom routers; switchers; and other network applications.

During recent years, another growing area of high-current applications includes large LED displays and LED lighting, with focus on backlighting, automotive, and general illumination applications.

The broad market portfolio usually includes building blocks for switching regulators and switching controllers with an input voltage of ~2 to ~100 V and a current of up to 10 A, LDOs of 0.2 to ~10 A, and a high, medium, or low precision band gap reference.

There are many other power analog and mixed signal applications. One of these applications is in the power over Ethernet (PoE). PoE uses existing Ethernet infrastructure to deliver power at a voltage level of 48–52 V, with a safety limit of ~60 V based upon the IEEE 802.3 standard. The maximum power is 30 W or greater. PoE combines power-sourcing equipment such as Ethernet routers, switches, and hubs at one end of the cable. At the other end is a powered device, such as a VoIP or, WLAN access point, network camera, or security system. In a typical application,

the power can be sourced from the power-sourcing equipment through the data lines via data transformer center taps.

One of the PoE challenges is the capability of the high-voltage mixed-signal integrated circuit with a fast digital signal to operate at 60 V. Other challenges include conversion efficiency, transient protection, ease-of-use/support, and the need for galvanic isolation.

Overall, the modern trends in this power product combine the following features: a high switching frequency, higher current levels, higher efficiency, and low power consumption during standby.

Adding to the value of this solution are enabling features such as power good, sync, sequence, tracking, and multiple outputs. Another advanced feature is digital control. Digital control allows for switchers with programmability and remote monitoring capabilities. The overall target is to design smaller, cheaper and faster switchers, as well as to provide high-voltage analog products (up to 100 V). The trends in low-voltage analog products are a higher output current (up to 20 A) and lower regulated output voltages.

Switching voltage regulators with a switching frequency of up to 50 MHz are currently under research for possible reduction of the form factor, using integrated inductors and capacitors or other integrated and co-packaged passives on the chip.

Unique performance characteristics can be provided by serial digital power control of analog and digital feedback loops, using data converters for digital feedback loop and monitoring.

7.1.1.2 ESD Challenges

Power analog product demands create significant challenges for on-chip ESD design (Table 7.1).

Table 7.1 ESD challenges generated by power market demands

Power market demands	ESD challenges
Low-cost solutions; smaller form factor solutions	Small footprint of ESD cells, under-the-bond pad ESD cells, self-protection
Fewest external components	New ESD circuits to explore on the product level
Highest efficiency	Low parasitic structures; ESD protection of very fast switching voltage nodes; low substrate noise
More current capability	Lower ESD protection window; power array protection, internal circuit ESD damage through the power array ESD signal transmission
Higher total output voltage accuracy	Precise ESD clamp characteristics
Dynamic control; faster response	Voltage reference ESD circuits with low dV/dt effect
Control and programmability; smart solutions	Multiple voltage domains; EEPROM protection, local snapback protection for digital and low-voltage control pins

To address the demands of the power market, a significant evolution of the integrated process technology is required.

The integrated components of the CMOS, BiCMOS, or BCD processes are designed toward increasing the switching speed. One of the major changes in physical parameters is the decrease of the $R_{DSON}C_{GS}$ product figure of merit, where R_{DSON} and C_{GS} are on-state resistance and gate–source capacitance of the integrated power component, respectively.

Reducing the form factor by integrating the typical external components on the chip, mainly the inductor and capacitor, allows for the design of very high frequency (>20 MHz) efficient switchers, presenting state-of-the-art solutions.

One of the major challenges of this solution is related to the power density of the integrated transistor arrays. On the technology level, this problem is addressed by the implementation of thick and thin copper-top metallization with low sheet resistance. Thus, other major figures of merit for integrated power arrays are the product of R_{DSON} and the area consumed by the power device ($R \times A$).

The trend toward control, programmability, and smart power circuits requires small signal circuit blocks. These blocks require an improved density with integration of digital and enhanced analog contents.

Many critical (life support) system applications, for example, medical and automotive, require a new level of ESD protection to guarantee a robust and reliable system performance, even in the case of overstress and upset events.

For portable applications, the small form factor and low cost require a compact, space-saving ESD design with under-bond pad solutions or solutions integrated in the power structure. Practically, integration of the ESD robustness into the process is required to avoid overdesign of the ESD structures. The usual expectation is that ESD clamp size is kept below the pad size. Structures that can be created in the process are expected to enable ESD clamps with dV/dt-triggered solutions (at least for switch, bootstrap, and power supply nodes).

In these conditions, the solution strategy is refocused on the protection of the individual internal circuit devices rather than the rails of the analog power circuit. This means, for example, that the high-side nodes are protected to the switch node, rather than to the ground, and only the low-side nodes are protected to the ground.

The new demands of analog power circuits also impact semiconductor process technologies. The most power-optimized processes include superior copper top metallization; high-density, low-voltage CMOS parts – down to 0.13 μm; an SOI process to eliminate substrate noise coupling; a superior laser trim capability (relative to Zener trim); power-optimized integrated high-current-density switch arrays.

These semiconductor process capabilities target low-cost power management units (PMU), digitally controlled switchers, and analog products with high voltage capabilities. They also provide voltage scaling capabilities.

The process development trend across industry generates more and more advanced and sophisticated BCD technologies under high cost-effectiveness.

Process technology is focused on the advanced drain–source on-state resistance R_{DSON} vs. the drain–source breakdown voltage, switching speed, as well as the elimination of latch-up effects at higher current conditions by effective deep trench isolation (10 μm and above), combined with high-energy buried layer and sinker implants.

A significant upgrade of the simulation tools is usually required before simulating the behavior of integrated switchers, taking into account the coupled substrate noise and ESD capability.

At high power density, constraints of integration of another technology become related to thermal packaging characteristics at higher power. Similar to discrete components, a substantial reduction of the thermal resistance is required.

Another packaging characteristic concerns finding a better way to produce high current out of the package with a small form factor. This involves advanced ball-bond and wire-bond technologies.

7.1.2 Integrated DC–DC Converters and Controllers

In this chapter, many application examples are related to switching voltage regulators. Therefore, this section provides some minimal introductory material on the topic. Among the DC–DC convertor product family, some of the most common products are the step-down buck and step-up boost switching blocks. However, other convertors, like buck-boost, multiphase, or resonant [122–124], can be designed with the same basic principles, even for producing a voltage of opposite polarity to the input voltage.

The most simple implementations of buck and boost converters using integrated CMOS power devices are shown in Fig. 7.1 for an asynchronous switching principle that involves the flyback diode and for a synchronous switching principle that involves two switching devices.

A buck converter (Fig. 7.1a, b) operates in continuous mode by using the current through the inductor supported by the repeatable commutation cycle. During the first phase, the MOSFET switch provides initial current through the inductor. Then, the energy stored in the inductor is transferred to the output of the converter and integrated in the constant voltage by the capacitor. After turn-off of the switch device, the current path is provided by the flyback diode. The total integrated output voltage is determined by the duty cycle controlled by the driver and varies linearly with the duty cycle for a given input voltage.

In an asynchronous buck converter, the freewheeling diode turns on after the switch turns off, as a result of the rising voltage across the diode. This voltage drop across the diode results into a corresponding power loss.

To overcome this loss, the synchronous buck converter is built with a modified version of the basic buck converter circuit topology, in which the diode function is performed by a second switching power array. In spite of the increased cost, the improved efficiency of this design is often preferable in portable applications.

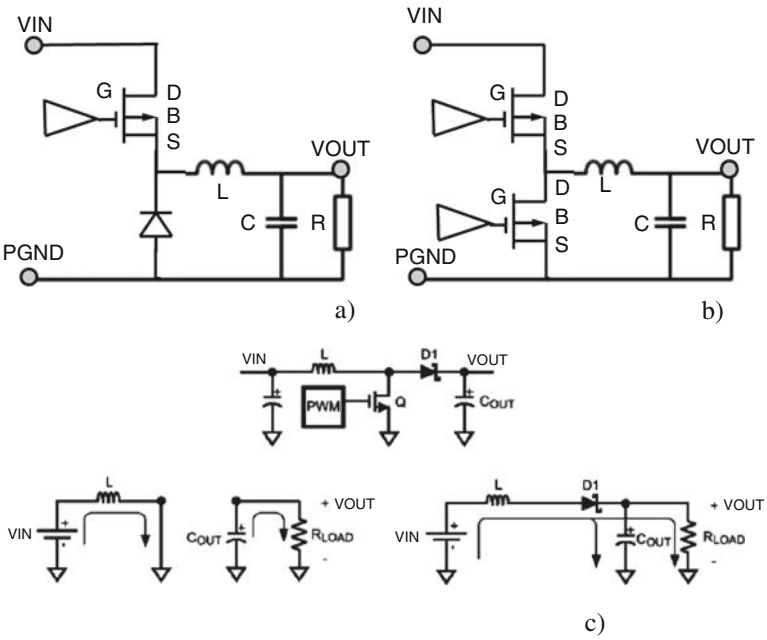


Fig. 7.1 Typical circuit diagrams explaining the operation principles for the step-down (buck) asynchronous (**a**) and synchronous (**b**) and step-up (boost) voltage regulator with diagrams illustrating principles of operation (**c**)

Another advantage of the synchronous converter is its bi-directionality. When power is transferred in the "reverse" direction, it provides a boost action that can be used to provide a power supply for the high-side driver.

A typical example of buck converter application is different battery chargers for portable electronic devices using a dc voltage supply, for example, an automotive on-board 12 V power outlet.

Alternatively, batteries, solar panels, or rectifiers can provide the power. To convert the original DC voltage level to a higher level, step-up or boost converters are used (Fig. 7.1c). For high efficiency, the switch must turn on and off quickly and have low losses. Therefore, MOSFET devices are usually chosen for highly efficient circuits.

Examples of application for boost circuits are portable LED lighting and hybrid cars, as well as cold cathode fluorescent tubes (CCFL) in devices such as LCD backlights.

The key principle that drives the boost converter is the inherent tendency of an inductor to resist changes in current. When being charged, the inductor acts as a load and absorbs energy, and when being discharged, it acts as an energy source. The voltage the inductor produces during the discharge phase is related to the rate of change of the current and not to the original charging voltage, thus allowing difference between input and output voltages.

In the on-state, the switch provides an increase in the inductor current. In the off-state, the switch is open and the only path offered to the inductor current is through the flyback diode D_1, the capacitor C_{OUT}, and the load R_{LOAD}. This path results in the transfer of the energy accumulated by the inductor during the on-state into the capacitor.

The boost converter may be operated in either the continuous or discontinuous conduction mode. Operation in the continuous mode is more efficient and provides lower EMI characteristics than the discontinuous mode. In continuous conduction mode (when the inductor current never reaches zero), the boost regulator operates in two cycles. In the first cycle of operation, MOSFET Q is turned on and energy is stored in the inductor. During this cycle, diode D_1 is reverse-biased and the load current is supplied by the output capacitor, C_{OUT}. In the second cycle, MOSFET Q is off and the diode is forward-biased. The energy stored in the inductor is transferred to the load and output capacitor. The ratio of these two cycles determines the output voltage.

A more detailed analysis and analytical description of converter operation can be found in [122–125].

Similar to buck converters, a synchronous version can provide better efficiency with the additional cost of another integrated NMOS array.

The typical ESD protection solutions for converter pins are discussed in detail in the following two sections, separately for low-voltage and high-voltage converters. From the circuit diagrams (Fig. 7.1), it is already expected that switch node protection will be a challenge, due to the fast transient voltages at this pin (connected to the inductor). Also challenging is the transient voltage at the input pin, due to the inductance of the bond wire at high current.

In general, DC–DC converters can be designed with different integrated or discrete components. Usually, high-quality factor passives are either external or integrated into the package. An alternative converter design may include integrated controller circuit driving the gates of discrete FET's, IGBT or GaN power devices. The ESD specifics of controllers will also be discussed in this chapter.

7.1.3 Integrated Power Arrays

7.1.3.1 Power Losses

One of the most intriguing, and complex topics related to analog power circuits is the topic of integrated power devices. The design of integrated power components greatly impacts the figures of merit of analog power products.

In the case of low-voltage power components, standard NMOS and PMOS are used. In the case of high-voltage circuits, integrated lateral double-diffusion MOS (LDMOS) (Fig. 7.2a) or vertical DMOS (VDMOS) (Fig. 7.2b) architecture is used for the integrated power device. A more advanced solution can be based on trench architecture, which are also used in discrete components (Fig. 7.2c). While the maximum operating voltage of the LDMOS devices is scalable by the lateral drift region,

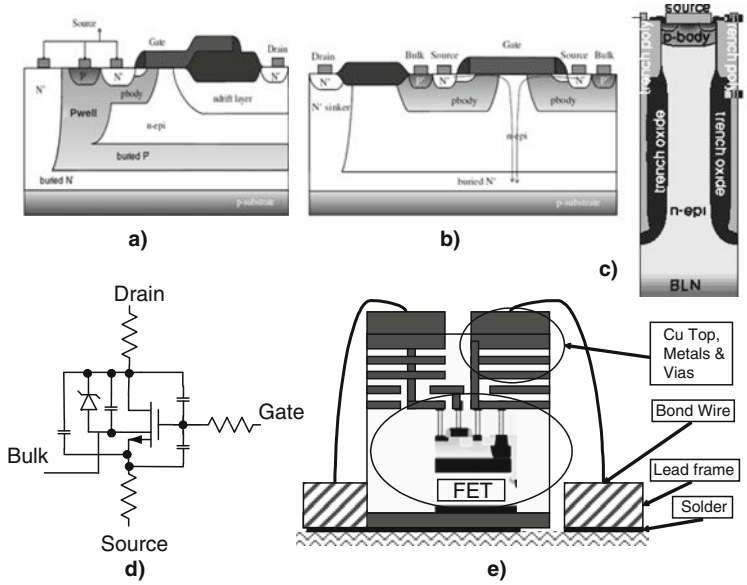


Fig. 7.2 Cross sections of RESURF lateral (**a**), vertical (**b**), and trench DMOS (**c**) devices; schematic representation of the integrated power component with parasitic components (**d**); and illustration of the resistive components in the final package (**e**) [126]

the vertical devices usually provide a fixed voltage level determined by the thickness and doping profile of the N-epi region.

The major goal in a switching circuits design is to increase the power efficiency of the circuit. There are five major loss channels associated with the switching device:

(i) The drain–source on-state resistance loss P_{RDSON}
(ii) The gate charge loss P_{GATE}
(iii) The switching losses P_{SW}
(iv) The deadtime losses P_{DT}
(v) The reverse recovery losses P_{RR}

The drain–source *on-state resistance* R_{DSON} is an equivalent resistance from drain to source. It simply leads to resistive-type losses in the switching circuit according to $P_{\mathrm{RDSON}} = I_{\mathrm{rms}}^2 R_{\mathrm{DSON}}$, where I_{rms} is the random mean square current in the circuit. In real applications, the resistive losses are accumulated across the entire network within switching current path. The total resistance is accumulated by resistive components from the print circuit board (PCB), solder, frame, bond wires, metallization layers, transistor, and inductor Fig. 7.2e:

$$R_{\mathrm{TOTAL}} = R_{\mathrm{board}} + R_{\mathrm{solder}} + R_{\mathrm{frame}} + R_{\mathrm{wire}} + R_{\mathrm{metals}} + R_{\mathrm{FET}} + R_{\mathrm{inductor}}$$

The *gate charge loss* Q_g represents the amount of charge necessary to raise the gate/source potential from the off state to a given voltage for on-state conditions (Fig. 7.3). In this case, the loss is proportional to the switching frequency f_{sw} and the voltage level provided by the external power supply or the internally regulated voltage of the gate driver V_{CC}; $P_{GATE} = Q_g\, V_{CC} f_{sw}$.

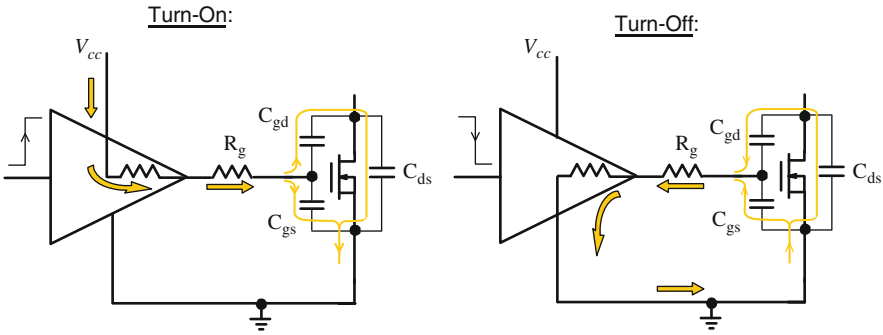


Fig. 7.3 Illustration of the gate charge losses in MOSFET during turn-on and turn-off [126]

Switching losses occur when the power FET is transitioning between the off- and on-states. The loss is affected by most of the parasitic resistances and capacitances associated with the device. The switching loss can be estimated according to the expression $P_{SW} = \frac{1}{2} f_{SW}\, V_{DS}\, I_D\, (t_r + t_f)$, where t_r and t_f are the corresponding times for turn-on and turn-off processes, respectively. Essentially, the switching loss occurs due to the finite time of overlap of the switch current and switch voltage waveforms. Respectively, the larger the gate charge Q_g, the longer the rise and fall time and thus the larger the switching loss component (Fig. 7.4).

The *deadtime loss* is related to the conduction process in the body diode during the deadtimes: $P_{DT} = f_{SW}\, V_{Diode} \cdot I_D\, (t_{DT1} + t_{DT2})$. During synchronous switcher deadtime, the power FET is off and the low-side body diode does not conduct. In an asynchronous switcher, the "deadtime" loss occurs during the entire duration of the off-state.

The *reverse recovery losses* are related to the charge Q_{RR} accumulated across the junction. This charge must be recovered when switching resumes (Fig. 7.5). In order to reverse bias, it is necessary to remove the accumulated injected charge in the depletion region of the diode. This is achieved by using current from the high-side MOSFET and creates the reverse recovery loss. Reverse recovery (RR) time can slow the turn-on transition and increase switching losses as well. Smaller deadtimes lead to smaller RR losses: $P_{RR} = f_{SW} \cdot \mathrm{VIN} \cdot Q_{RR}$.

The peak efficiency point is primarily controlled by the on-state resistance R_{DSON} and Q_g. Thus, the integrated or discrete FETs inside (or off-chip) a switcher determine the *peak efficiency* point (Fig. 7.6).

A major consideration is a trade-off between product performance and manufacturing cost. The cost increases with the number of smaller features in the design; the number of mask and process layers; die size; Cu top usage; package type; multi

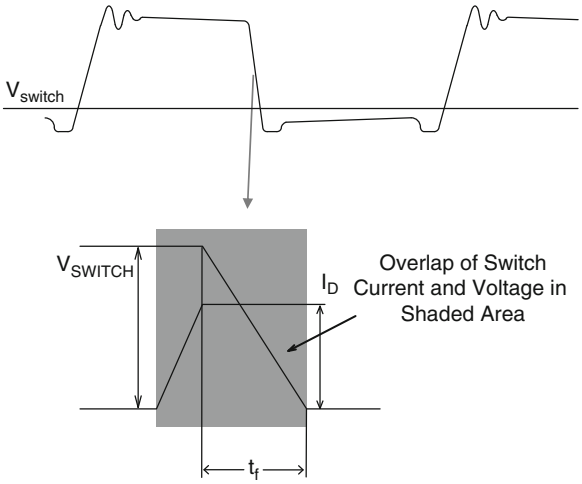


Fig. 7.4 Illustration diagrams for the MOSFET switching losses [126]

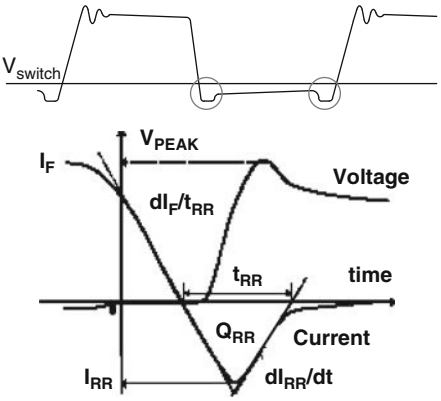


Fig. 7.5 MOSFET losses – deadtime and reverse recovery [126]

die vs. single die usage; die attachment steps; number of bond wires; lead frame, uSMD, and material vs. process costs.

Thus, efforts are focused not only on reducing the resistance in the mean current path to the lowest square count possible in the metal, but also on avoiding current crowding and being able to bond to the package.

With increase in the total power array area, the relative contribution of the channel resistance is reduced and metallization resistance become dominant (Fig. 7.7a). There are several layout techniques that are mainly empirically developed. For example, the "interdigitated" or so-called "Christmas tree" metal pattern helps to avoid current crowding (Fig. 7.7b). The main idea of this structure is to roughly equalize the resistance to any point within the FET. Sometimes, current density may

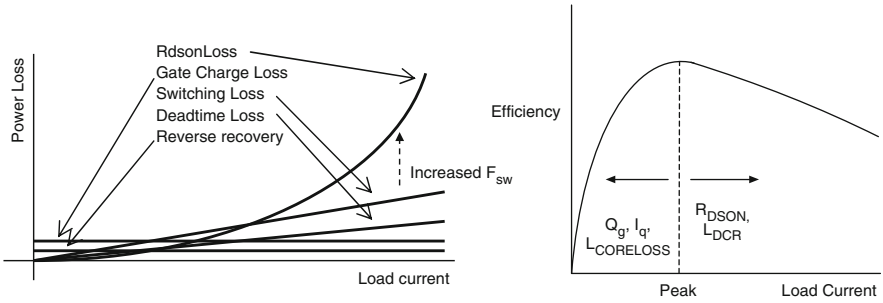


Fig. 7.6 Losses over output current [126]

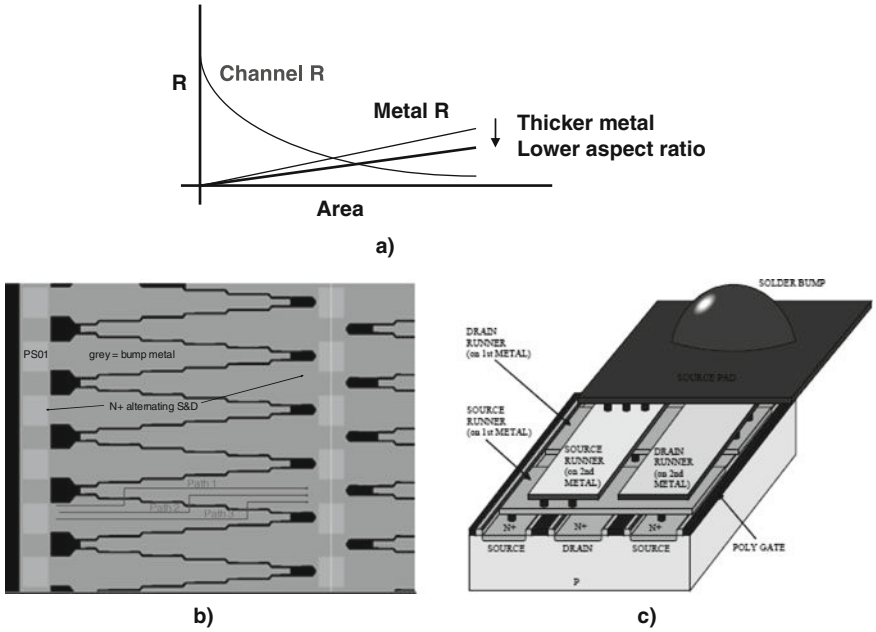


Fig. 7.7 Dependence of the array resistance components upon area (**a**), example of the "interdigitated" copper top (**b**); and three-metal micro-SMD layout for power FET array (**c**) [126]

require parallel layers. This approach might be especially effective with a copper top metal layers.

One of the major trade-offs for R_{DSON} is the reduction of the breakdown voltages. Data across industry analog processes are summarized by feature size in Fig. 7.8 for published sub-50 V processes. One of the major problems is related to the ESD protection capability of the particular design of a power array. This problem is brought into focus in the following section.

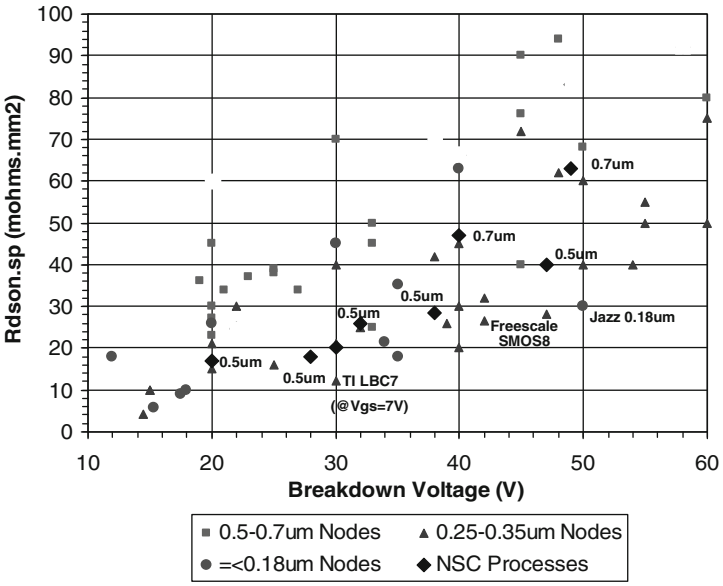


Fig. 7.8 Dependence of the on-state resistance per unit area vs. breakdown voltage across industry analog processes derived from published data

7.1.3.2 Self-Protection Capability (SPC) of Integrated Power Arrays

According to the previous section, power arrays are primarily optimized to minimize losses. This, however, does not automatically guarantee a suitable performance of the arrays in ESD conditions [127, 128]. The aspects of array self-protection capability (SPC) assessment by experimental methods are the focus of this section, with the simulation methods described in the following section.

It is logical to assume that the self-protection capability of the power array is determined by two coupled factors: device-level effects, governed by the device design, and array-level effects, mainly determined by the distributed current conduction across the array surface from the pads. Device regions and doping profiles determine the critical avalanche current per micron width, on-state parameters, and gate coupling. The array layout design is responsible for the non-uniform current conduction of the array due to both the distributed interconnect resistance and the gate network.

In principle, in power analog ICs with integrated power arrays, a major challenge is the protection of the IC pins connected to the power array. Most typical examples of such pins include the switch pin in DC–DC converters, the gate pin in controllers, and the output pins in LED drivers. Due to high transient voltages, an ESD clamp cannot be used and the ESD protection network typically relies on the SPC of the power array itself.

At the same time, for an array voltage above 20 V, the snapback event usually results in immediate device failure [129], with exception of some rare cases [130]. This phenomenon is observed even in very large integrated power arrays.

An example of failure in a large integrated 100 V power array with 60 mm total gate width is presented in Fig. 7.9. The array is a part of an asynchronous buck switching voltage regulator circuit. The local burnout of the array is observed at a 2 kV HBM stress. If calculated with assumption of uniform current distribution across the array, the failure corresponds to an average current density of only 22 μA per micron width.

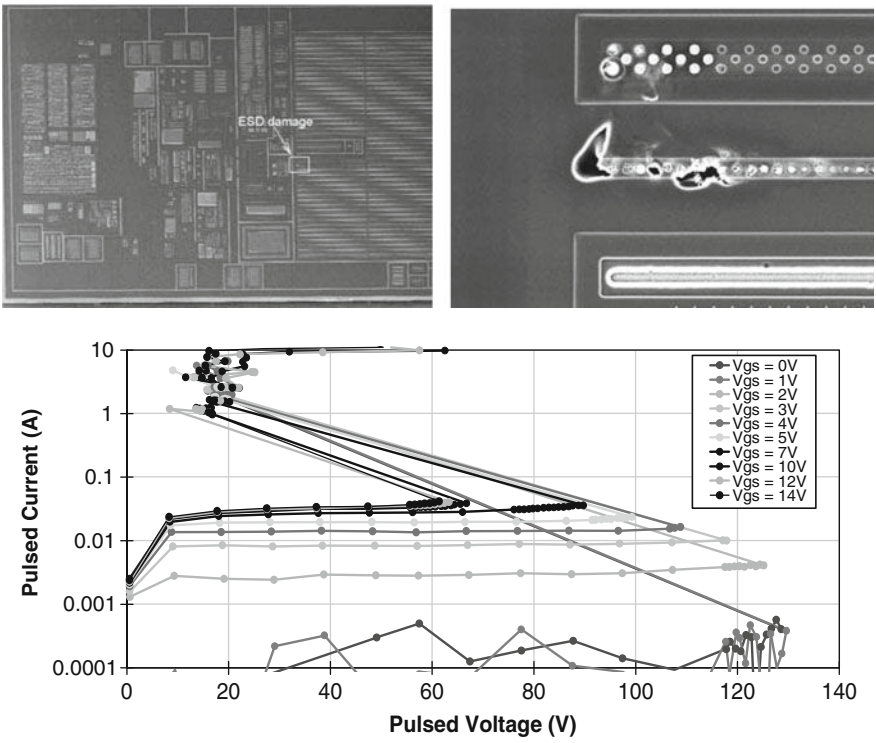


Fig. 7.9 Local $W = 60$ mm array damage in power IC case and 100 ns pulsed characteristics of the output array for $W = 1$ mm

As it was described in the section above, integrated analog power products are usually designed with significant trade-off between the on-state resistance and the breakdown voltage. As a result of such aggressive optimization, the SPC of an array in the ESD pulse regime is primarily determined by the amount of avalanche current level that the array can withstand prior to entering the irreversible snapback mode. This figure of merit is somewhat similar to the avalanche energy listed in the data sheets of discrete MOSFET and IGBT components (Section 8.5).

Several components of successful power array design include (i) an experimental methodology for evaluation of the self-protection capability of integrated power arrays in ESD pulse regimes and (ii) design methods to improve the SPC or the NLDMOS arrays on the device and array levels.

It is important to emphasize that the focus of this section is on improving the SPC of large integrated arrays by increasing the current in the critical regime before snapback rather than attempting to achieve a reversible operation of the NLDMOS device in the snapback mode [129], in opposite to the examples described in Section 4.5. In general cases, attempts to achieve reversible operation in snapback mode are usually unsuccessful [130].

In particular, a significant increase in the SPC of the array in *non-snapback mode* can be achieved by avalanche current balancing using a rather deep N-sinker implant in the drain-diffusion region. A practical methodology for array comparison is proposed and experimentally validated in [131].

A comparative analysis of the arrays with different designs is presented below. All measurements presented have been carried out on wafer level. For comparative analysis of the DC electrical parameters of the power arrays, an industrial HP electrical tester has been used to obtain precise array on-state resistance (for drain–source voltage of 0.1 V and 5 V gate bias) and breakdown voltage (at 1 μA drain current).

The self-protection capability is characterized as the dependence of the critical HBM failure level on constant gate bias (Fig. 7.10a, b). Additionally, transmission line pulse (TLP) measurements have been used in a similar setup (Fig. 7.10a) to obtain the 10 ns rise time pulsed SOA of the arrays and to establish the dependence of the snapback triggering point (I_{T1} and V_{T1}) on gate bias, as in Fig. 7.9.

The experiments presented below investigate the effects of drain contact region changes and alternative source region design on the self-protection capability of 1 and 24 mm total gate width NLDMOS arrays.

The layout of the original design of the NLDMOS array is presented in Fig. 7.10c. This design combines an n^+-drain region with an n-drift implant and has a polygate region enclosing the source n^+-region and the p^+ and Pbody regions. Using conventional NLDMOS design practice, the second diffusion Pbody is aligned by a polymask to form the body of the MOS structure.

The critical avalanche current of the NLDMOS is determined by the parasitic n–p–n structure. The degrees of freedom for improvement of this avalanche current level are limited by two major constraints: (i) improve the array SPC without process change, i.e., using only available mask layers (for example, varying the Pbody implant dose to reduce the internal base resistance of the parasitic n–p–n, which has been optimized for the operational regime of the NLDMOS, is not allowed); (ii) deliver a solution that provides on-state resistance and other relevant figures of merit that meet or exceed those of the reference design (Fig. 7.10c).

Three distinct and appropriate methods of increasing SPC have been incorporated into experiments based on related studies [129, 130, 132–134]:

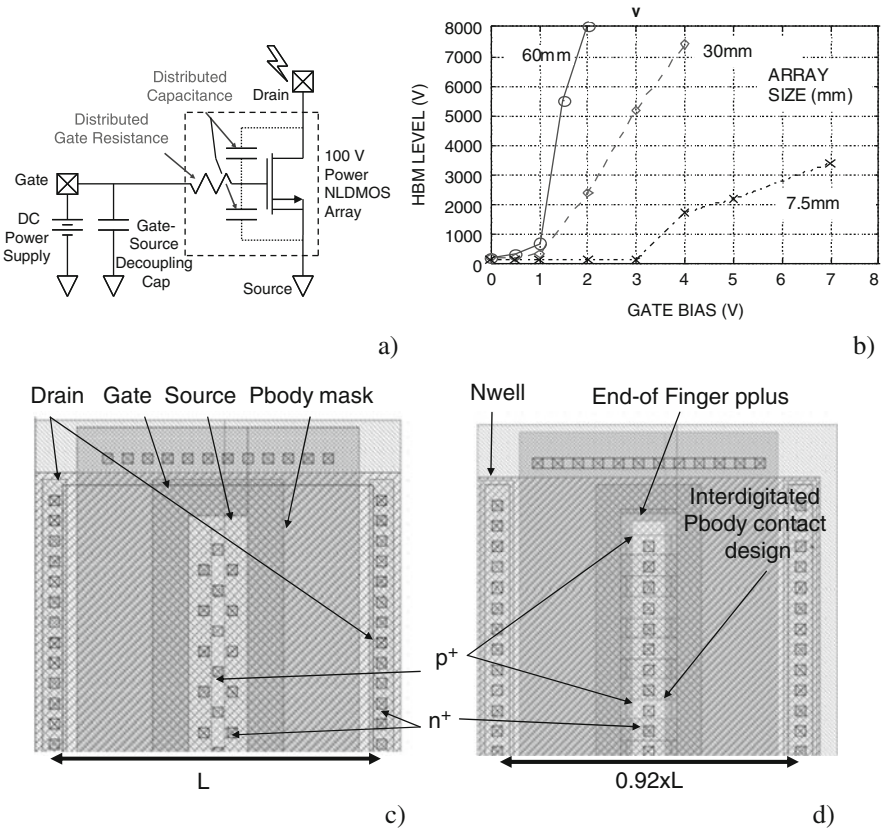


Fig. 7.10 Experimental setup for SPC evaluation (**a**); measured dependence of the critical HBM level upon the dc gate bias for different array sizes (**b**); snapshot layout view of the conventional 100 V NLDMOS array (**c**); and the array with modification of several design parameters including 8% reduction in size (**d**)

The first alternative device design implements interdigitated n^+-source/p^+-Pbody diffusion regions (Fig. 7.10d). This approach has been previously used [4, 5] primarily as a measure to achieve reversible snapback operation. Interdigitation ratios range from every other contact to only six p^+-regions along the 100 μm device fingers. This approach brings the added benefit of reduced source region layout area. A chief concern with this approach is possible impact on R_{DSON}.

The second device design measure includes blocking the n^+-implant at the ends of the fingers (Fig. 7.10d). A drawback of this approach is the slight R_{DSON} loss due to reduction in the source layout area in arrays with short fingers.

The third measure includes additional Nwell or Nsinker implants in the drain contact region. This measure has been discussed as an improvement that may provide reversible snapback operation.

For each process technology it is important to understand and quantify the effects of these measures on SPC improvement.

In analog circuit applications, different conditions can be realized at the gate of a power NLDMOS array, so comparison of the self-protection capabilities of power arrays requires gate control of the arrays. In this study the dependence of the critical HBM level (the highest voltage HBM discharge that the array safely withstands) upon gate bias is treated as the key SPC figure of merit.

According to the SPC dependence (Fig. 7.10b) the total HBM current through the array is a superposition of two components, channel and avalanche current. The channel current depends on the DC gate bias and the gate coupling during a transient HBM event. As gate bias increases, an increasing fraction of the HBM current is safely conducted through the channel, leading to the observed increase in HBM passing level.

The critical avalanche current is limited by the electrical avalanche–injection current instability, which is caused by the parasitic n–p–n BJT structure in the NLDMOS device. In the avalanche–injection regime, conductivity modulation under the critical current density is controlled by the parameters responsible for current gain and avalanche multiplication in the parasitic n–p–n structure.

The major results of the 1 mm array studies at 0 and 5 V gate bias are summarized in Fig. 7.11 and Table 7.2, in comparison of the original reference NLDMOS device A01 with improved devices A16, A17, and A21. The best device, including all three features described above, is compared to the reference device in Fig. 7.10c. These measures have increased the critical avalanche current at zero gate bias by two orders of magnitude, and the snapback voltage has been increased by 15–20 V at both 0 and 5 V gate bias (Fig. 7.11a).

The following conclusions can be drawn based on the analysis of the experimental results (Table 7.2). The R_{DSON} loss due to interdigitation of the source–Pbody is insignificant. For the minimum n^+-source area, when n^+-contacts are only 50% of the source region area (alternating with p^+-diffusions), R_{DSON} increases only 2.6%, which is compensated by a 7.9% R_{DSON} improvement due to the smaller cell dimension. For the device with only a 1:9 p^+/Pbody to n^+-source area ratio, the R_{DSON} increase due to the reduced source area is only 1%, with a total R_{DSON} improvement of ~7%. However, interdigitation of the source and Pbody does not by itself provide any SPC advantage. The critical current of these arrays is similar to that of the reference array.

Both the drain Nwell implant and the end of finger p^+-implant lead to at least one order of magnitude improvement in the critical avalanche current.

For the array already upgraded with the drain Nwell implant, the end of finger p^+-implant improves the avalanche current level by a factor of 2.

These various methods also lead to an increase in the snapback voltage of the array at the gate bias of 5 V.

These results and conclusions for 1 mm arrays are confirmed by results from 24 mm arrays with the same design alterations.

The effects of the end of finger p^+-diffusion and the inclusion of Nwell or Nsinker in the drain have been analyzed and supported by numerical simulation. An N-sinker

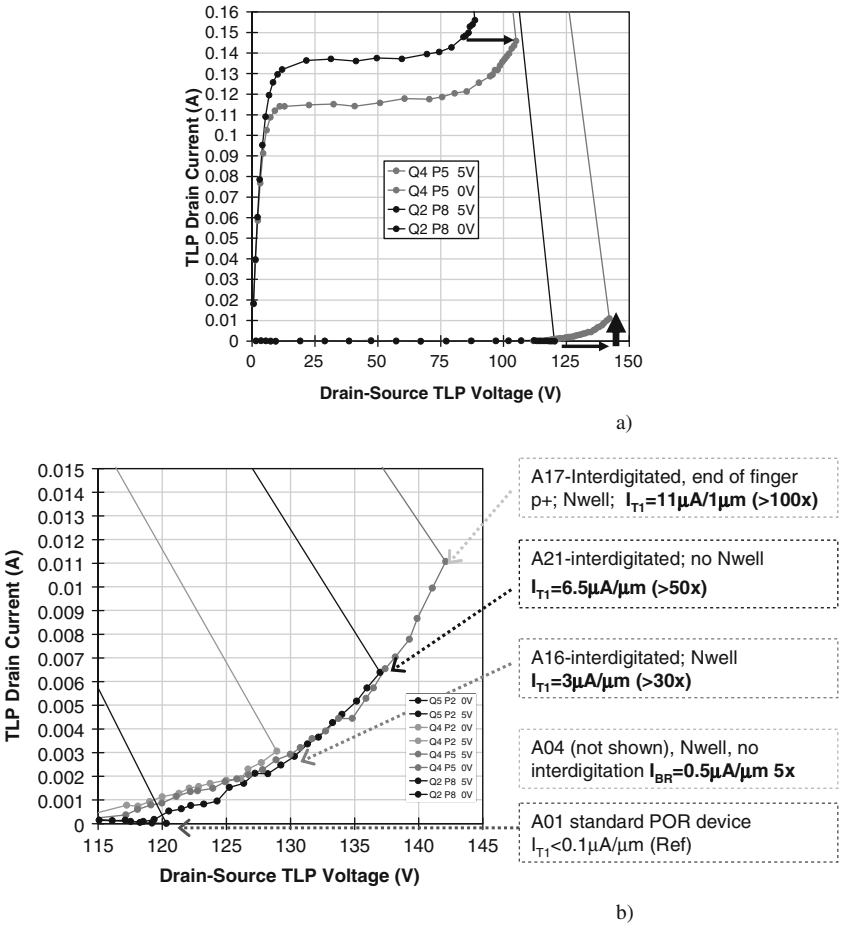


Fig. 7.11 The effect of different array design features. Comparison of the experimental TLP I_D–V_{DS} characteristics for the standard NLDMOS and the devices with different design changes

implant provides significant improvement of the critical avalanche current level, but requires a substantial increase in the drain length of several microns to avoid breakdown voltage reduction and is thus impractical.

As expected the major effects can be understood using avalanche–injection conductivity modulation in n–p–n structure with corresponding critical conditions $\alpha M > 1$, where α is the current gain and M is multiplication coefficient. The Nwell implant is too light to have a significant effect on current ballasting when implanted at the drain. According to 2D TCAD analysis (Fig. 7.12b), the benefit of drain Nwell implant is reduction of the avalanche multiplication factor M, resulting in achievement of the instability criterion at a higher critical voltage and current, thus providing the SOA improvement.

Table 7.2 Pulsed and DC characteristics for reference and improved devices

1 mm cell	Features description	$V_{gs}=0$ V_{T1} (V)	$V_{gs}=0$ I_{T1} (mA)	$V_{gs}=5$ V V_{T1} (V)	$V_{gs}=5$ V I_{T1} (mA)	V_{br} (V)	R_{DSON} × area
A01	Reference	120	0.018	88.6	156	117.4	363
A16	Nwell, interdigitated with nine p^+ islands	129	3.05	102	148	108.3	322
A17	Nwell, interdigitated with nine p^+ islands and p^+ end of fingers	142	11.1	105	146	108.7	339
A21	No Nwell, interdigitated with nine p^+ islands and p^+ end of fingers	137	6.39	97.3	140	111.8	359

Understanding the second dominant effect requires taking into account possible manufacturing effects resulting in the formation of spherical (as opposed to cylindrical) junctions in the corners of the poly ring due to the tilted implant shading. As a result the base of the parasitic n–p–n structure in the corners becomes lightly doped, thus providing elevated gain α and enabling the avalanche–injection current instability in the device at low critical currents. The doping profile obtained from 3D process simulation shown in Fig. 7.12c demonstrates surface reduction of the developed doping level in the corners of the polygate.

Thus a significant increase in the SPC of the array in non-snapback mode can be achieved by both changing the avalanche multiplication coefficient at the drain region using an Nwell implant and eliminating the cylindrical junction in the corners of the source region. This improvement is achieved using "free" approaches that do not require process alteration or result in degradation of the major DC and switching characteristics of the power arrays.

7.1.3.3 Physical Simulation of DeMOS Power Arrays in ESD Regime

A simulation methodology [135] applied to study the physical effects in large high-voltage power arrays demonstrates that the critical HBM pulse regime for snapback in a power array is a non-linear function of the gate bias. This simulation has also been used to visualize the scenario of local burnout of the power array as a result of localized filament formation. Analysis of array operation under ESD stress is complicated by the unpredictable effect of the gate driver circuit on gate coupling. The physical processes in ESD conditions in such a large distributed object as a power array are not well understood due to the grid size limitations of conventional simulation tools.

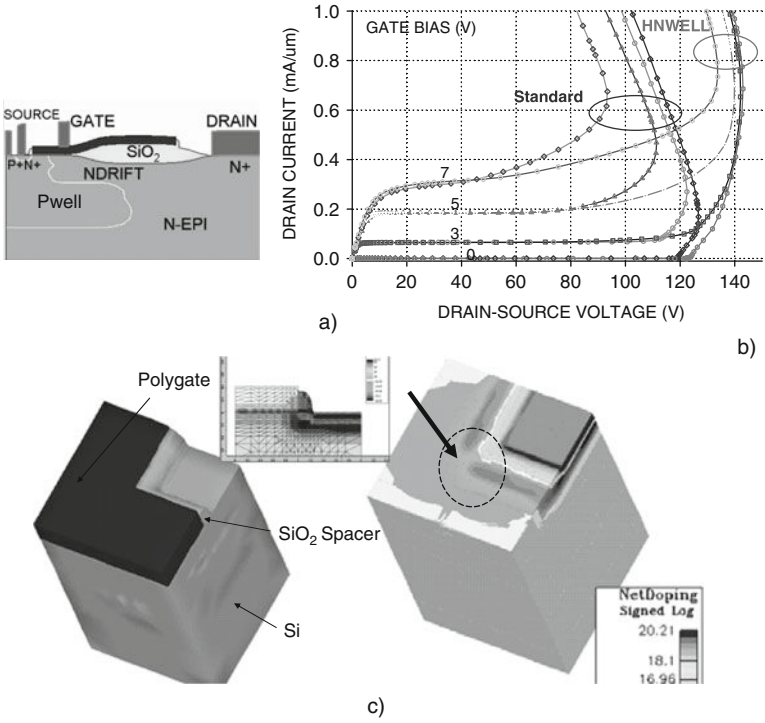


Fig. 7.12 A 100 V NLDMOS TCAD cross section (**a**) and comparison of TCAD I_D–V_{DS} characteristics for the standard NLDMOS and the device with additional HNWELL implant (**b**); (**c**) 3D structure of the NLDMOS gate–source segment and simulated doping distribution produced by implants (P-body, NLDD, n^+) after all diffusions with dose, energy, tilt and rotation from 2D calibrated process flow

The methodology for analysis of snapback behavior and current distributions in power cells and arrays under ESD stress has been first proposed and experimentally validated in [135]. It includes extraction of the array netlist using the interconnect clustering method, construction of the equivalent netlist using compact snapback device models (Fig. 7.13), and circuit simulation and visualization of the current distribution. This methodology enables mapping of the current distributions across the surface of the array in order to visualize the current distribution on transient timescales. The specific goal of the simulation is to understand the physical processes responsible for distributed power array burnout under ESD stress. Successful modeling of the non-uniform current distributions in large arrays may assist optimization of the cell and array topologies and the safe operation area in the ESD regime.

A 2.5D simulation of a 24 mm power NDeMOS array has been carried out across a range of constant gate biases. The simulated waveforms produced by the array under stepped increases of the HBM pulse have been monitored and recorded. The

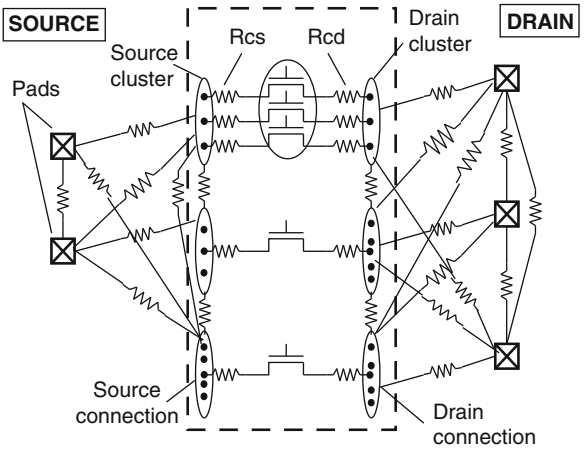


Fig. 7.13 Extracted distributed RC model of the power array

point of snapback is considered to correspond to the critical regime that leads to real array burnout.

Examples of the simulation results are shown below. At zero gate bias, reversible non-snapback operation is possible only up to the low HBM level of ~80 V due to a finite avalanche current and coupling of the non-silicided gate of the array. At a critical voltage level near ~90 V, however, a significant voltage drop and current increase indicate snapback of the array (Fig. 7.14a for 90 V HBM). For higher HBM amplitudes, the delay time for snapback is reduced (Fig. 7.14a for 1 kV HBM).

One of the important simulation results [135] is that the current distribution remains uniform both before and after snapback (Fig. 7.14c). Only after certain delay is a highly non-uniform state formed in the array (Fig. 7.14d). This non-uniform state corresponds to the current filament solution and is observed until the end of the ESD pulse.

When a DC gate bias is applied, the passing level of the array increases due to the addition of a channel current to the avalanche current. The dependence of the passing level upon the gate bias is presented in Fig. 7.15. Depending on the gate bias conditions, the filament can be spontaneously formed in two opposite corners of the array.

A two-pulse scenario for physical failure of the integrated power array is proposed based upon the experimental and simulation results presented above [135]. Reversible operation is possible until the critical pulse regime, as long as the snapback mode is not realized. *At first critical pulse* a uniform turn-on with negative differential resistance is followed by a non-uniform turn-off with a local filament. This results in local melting or accelerated electromigration until the end of the first pulse and formation of the array with local nonuniformly distributed damage. After zap leakage changes only at a voltage close to the breakdown level. *At the*

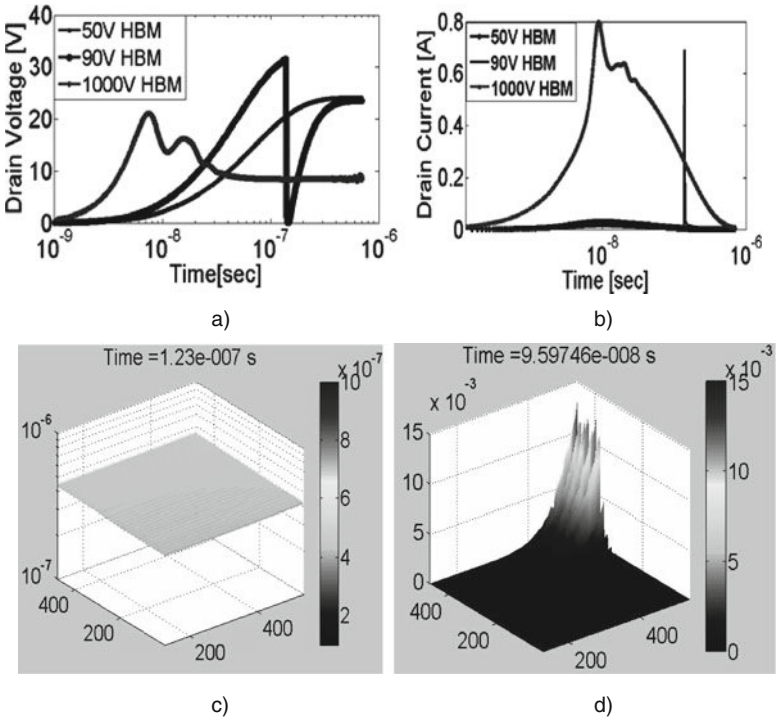


Fig. 7.14 Simulated voltage (**a**) and current (**b**) waveforms for 24 mm NDeMOS array at different HBM pulse levels in zero gate bias conditions and mapped simulated array current distributions for 90 V HBM pulse before breakdown, (**c**) and after snapback (**d**)

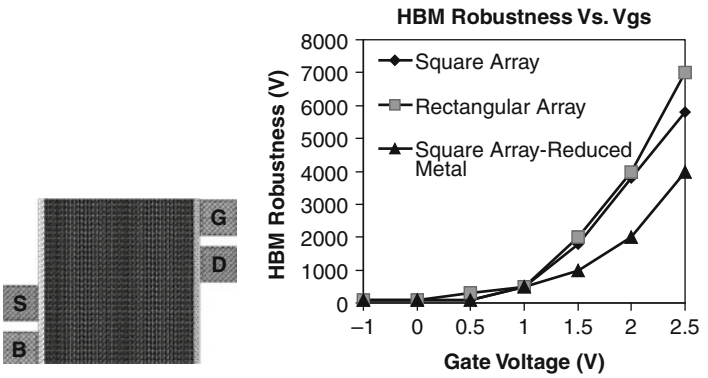


Fig. 7.15 Example of the layout view for a square 24 mm array and simulated critical HBM pulse level as a function of the gate bias

second critical pulse a local turn-on array in the pre-damaged spot is followed by an immediate local burnout due to current channeling and creation of the high leakage path.

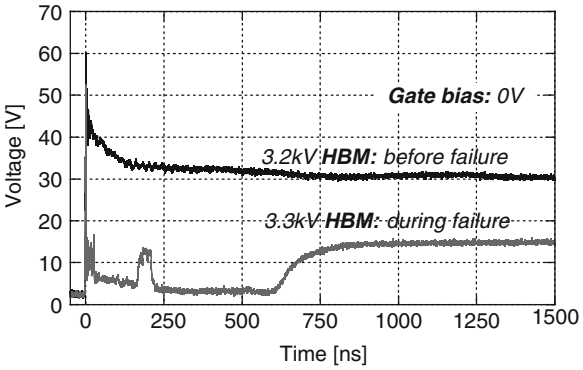


Fig. 7.16 HBM voltage waveforms before and during failure measured for 24 mm 20 V NDeMOS array

This conclusion correlates with the experimental data presented in Fig. 7.16, where the HBM waveforms for the array before and during failure are presented. Near the critical regime at 3.2 kV HBM, the array fully passes the test. However, at the critical pulse level, the change in the waveforms indicates the failure at the beginning of the pulse. Since the current distribution is uniform according to the simulation analysis at the beginning of the pulse, it is logical to assume that the array already receives local irreversible changes after the first pulse due to local current filament, as described above.

7.2 Low-Voltage Power Circuit ESD Cases

7.2.1 LV Power Switching Blocks

Similar to signal path analog circuits, an ESD protection network for the low-voltage power switching blocks can be realized with both with local snapback and rail-based active clamp protection.

In the case of rail-based active clamp protection, the design has a relatively low risk of missing the ESD protection specification. In this case, protection of the open drain pins is significantly more predictable, since the active clamp provides a much lower voltage waveform during ESD stress. However, this approach is not always compatible with hot plug-in requirements and conditions with fast power supply ramp voltages.

An example of a 5 V power train with active clamp protection is illustrated in Fig. 7.17a. In this low-risk design, the 5 V active clamp cells protect both the control and power pins. Respectively, each protected pin includes isolated Nwell and Pwell diodes connected to the ESDP and ESDM buses. The diodes provide an ESD current path that depends on the pin-to-pin combination.

The ESD protection network for the whole power train domain uses a common RC-timer module to provide a corresponding turn-off signal for the distributed

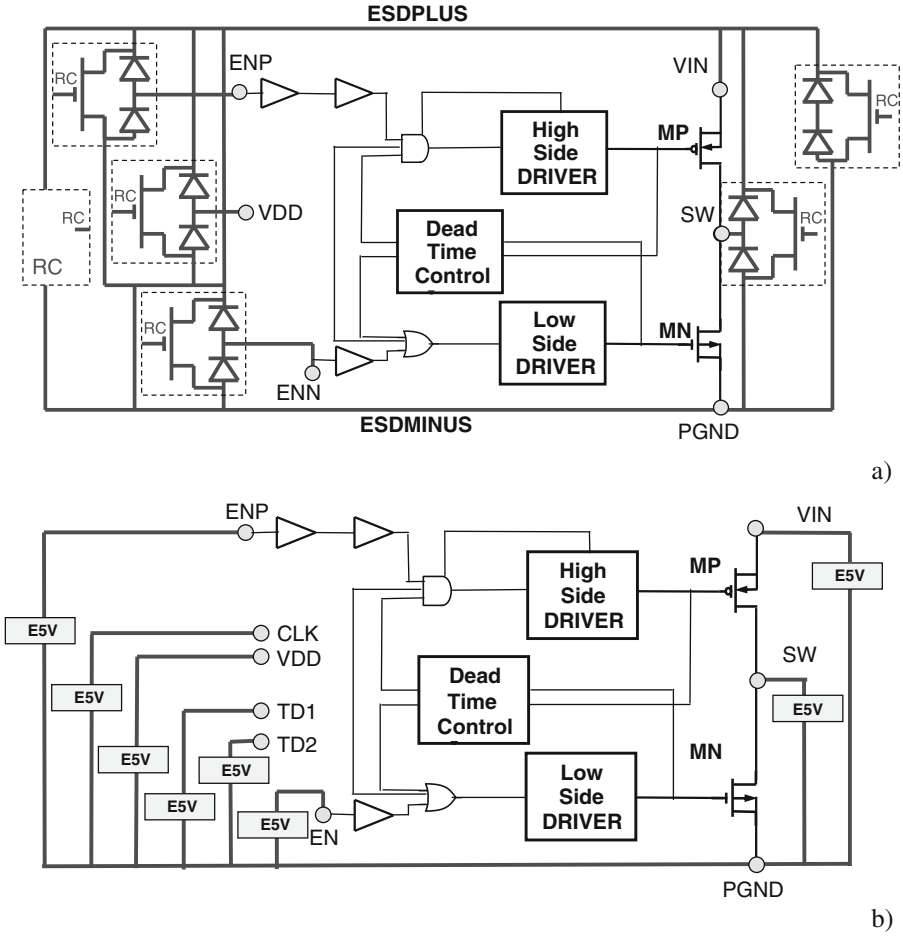


Fig. 7.17 Low-voltage power train protection using the rail-based active clamp approach (**a**) (not all control pins of the network are shown) and the same circuit protection using 5 V snapback NMOS clamps (**b**)

active clamp cells. The created network includes the RC-timer cell, upper and lower ESD diodes, and active clamps based upon a 5 V NMOS device (Chapter 4).

The design requires metal ESDP and ESDM buses to provide a low resistive path.

In sufficiently large switch devices (such as PMOS MP and NMOS MN devices), the switch pin SW does not require explicit ESD diodes, since they are already represented in the ESD network by the body diodes of the MOS devices. However, an appropriate metallization for the body diodes should be implemented.

Depending on the circuit specifications, the active clamp circuit network may be inapplicable, for example, in operation regimes with a fast variable input voltage (VIN). Therefore, an alternative, more aggressive design could use 5 V snapback NMOS solutions at every protected pin (Fig. 7.17b).

Typical control pins are enable, clock, and delay signal pins. They can usually easily tolerate an additional second-stage resistor of ~1 kΩ and thus typically do not create ESD protection challenges.

A low-voltage avalanche diode with a breakdown voltage of ~7 V can be advantageous if the process provides an opportunity for its inclusion. The diode can be effectively used to realize a two-stage protection clamp (Chapter 4) for the control pins.

Depending on the SOA of the standard device and the circuit design, the external VDD power supply for the driver can also tolerate snapback ESD protection.

Thus, a major challenge in the case of LV power train is the local snapback protection of the switch pin SW. In general, the challenge may include both the protection of the output NMOS array MN and the driver circuit connected to the NM gate.

If the snapback NMOS clamp "E5V" (Fig. 7.17b) is based on the same device, the ESD current path will depend on the conditions of the MN gate during ESD pulse, the well tap diffusion to source space, as well as the size and layout of the MN device. Since switch pin SW operates on the inductive load, fast transient voltages are expected at the pin. In this case, a simple grounded gate NMOS solution that utilizes the dV/dt is inapplicable due to gate coupling.

For a process with an insufficient ESD protection window due to the low SOA of NMOS and PMOS devices, there are no universal recipes for local protection of the SW pin and the final solution usually depends on both the process and product specifics.

If the process provides low-voltage avalanche diodes, the reference voltage in the drain–gate circuit can provide low dV/dt clamps with a triggering voltage in the ESD protection window (Section 4.3.1). In large arrays, a self-protection capability can be realized at the switch pin with appropriate layout and driver design.

The process-specific approach relies first of all on the pulsed SOA of the devices and the availability of the avalanche breakdown voltage reference. A simple demand of the process technology would be a requirement to increase the SOA limits for a 5 V NMOS. However, this conflicts with the power-optimized integrated components required to achieve superior switching characteristics and efficiency.

Recently, alternative solutions that involve active ESD clamping by the output array itself have been demonstrated through sophisticated driver circuit modification. However, these solutions are very product-specific.

7.2.2 Step-Down DC–DC Converters

ESD protection for complete low-voltage DC–DC converter products is similar in principle to the protection of the power switching blocks discussed above. However, the particular analog product specific implies additional constraints due to the size of the output arrays, pin signals, absolute maximum limits of the product, as well as the additional requirement for the ESD protection network to include other

circuit modules. Digital and analog blocks with different voltage domains, EEPROM memory, and other devices can represent these modules.

An example of a step-down 6 MHz DC–DC converter with a maximum load current of ~0.5 A is presented in Fig. 7.18.

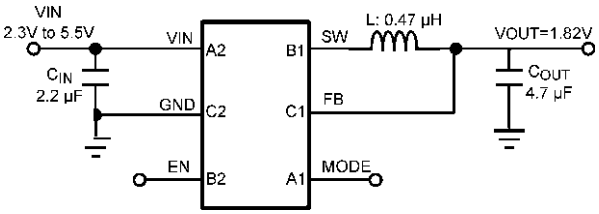


Fig. 7.18 A typical application circuit for fixed output voltage

The final, small form factor product is implemented with copper top metallization in a micro-SMD package. The last also enables ESD cell placement under the micro-SMD bumps. The under-the-pad design requirement often limits ESD clamp implementation to using only one of the metal layers.

The peculiarity of this particular example is the internal CBOOT pin with an internal bootstrap capacitor (Fig. 7.19) that eliminates the need for package-level protection in this pin. As a result, the ESD protection network is rather simple. The noisy input voltage (VIN) node is protected by an active clamp. The control pins MODE, EN are protected by local clamps with a two-stage isolation resistor to the internal nodes.

The most sophisticated switch node SW cannot tolerate a snapback device solution due to the fast transient voltages and either relies on the self-protection capability of the array or uses the same active clamp network as the VIN pin.

In this particular case, required ESD performance is achieved using the following approach. The ground pin GND is at zero voltage and connected to the global ESD

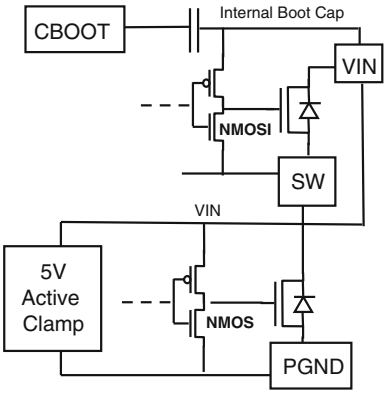


Fig. 7.19 Active clamp ESD protection of the VIN pin with simplified representation of the power side with an internal CBOOT pin and integrated boot capacitor as an internal node DeepNwell is tied to CBOOT

bus. The VIN is protected by the 5 V-tolerant active clamp. The 5 V switch pin SW is left without an explicit ESD clamp, relying on the self-protection capability of the relatively large output arrays. The remaining control pins are protected by a two-stage snapback NMOS clamp with a 1 kΩ two-stage resistor.

Such a mixed clamp approach is implemented to provide substantial noise isolation between the analog and power domains. It combines a snapback clamp at the control pins and an active clamp at the power pins. This solution also introduces diode isolation for the internal analog and power grounds, thus reducing the coupling between the domains.

The body diodes of the output NMOS and PMOS arrays, for a 0.5 A output current used in the synchronous output stage, provide a sufficiently high current path to the active clamp rails (Fig. 7.19).

Another essential part of the ESD protection approach is the local active clamp design. For a small-pin-count product, such an active clamp needs to be designed as a local clamp rather than a distributed network.

Measures should be taken to avoid false snapback during a fast transient at the VIN pin. Fast transients are produced due to the finite bond wire inductance. For example, the clamp can be designed with a fully butted NMOS source/well to prevent this problem. Additional attention should be paid to the metallization routing. At the same time, rise time control should also be implemented on the trigger line to avoid unexpected turn-on. An oversized NMOS device in the active clamp with a total width of $W = 10$ mm is comparable to the size the array of $W = 15$ mm connected to SW pin.

Alternative ESD protection of the switch pin by the local clamp and the related specific is discussed in the next section.

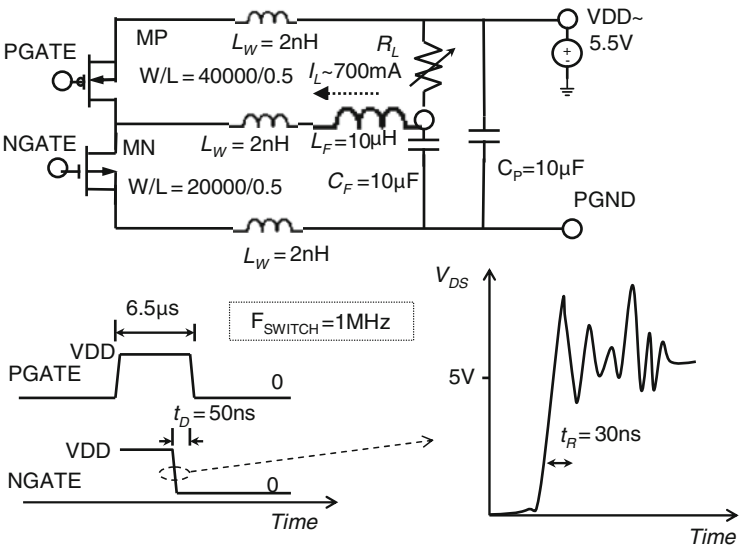


Fig. 7.20 Simplified output circuit block with external components, and waveforms for the driver signals at the gate of the output NMOS and PMOS arrays

7.2.3 Local Snapback Protection of LV Switch Pin

7.2.3.1 Case Study

A possible parasitic effect of the snapback NMOS grounded gate clamp connected to the switch pin is the detection of the transient current due to gate coupling above the threshold voltage. This effect usually precedes a false triggering of the clamp in the transient latch-up mode.

Thus, at some critical level of the input voltage, ESD clamp snapback may result in damage of the output circuit and switch devices. This effect is illustrated below. For this particular product case, local burnout of the switching power NMOS has been observed for circuit operation at elevated input voltage levels (VDD).

A simplified output circuit block diagram with external filter components $L_F C_F$, decoupling power supply capacitor C_P, bond wire inductance L_W, and driver signals at the gate of the output NMOS and PMOS arrays is presented in Fig. 7.20.

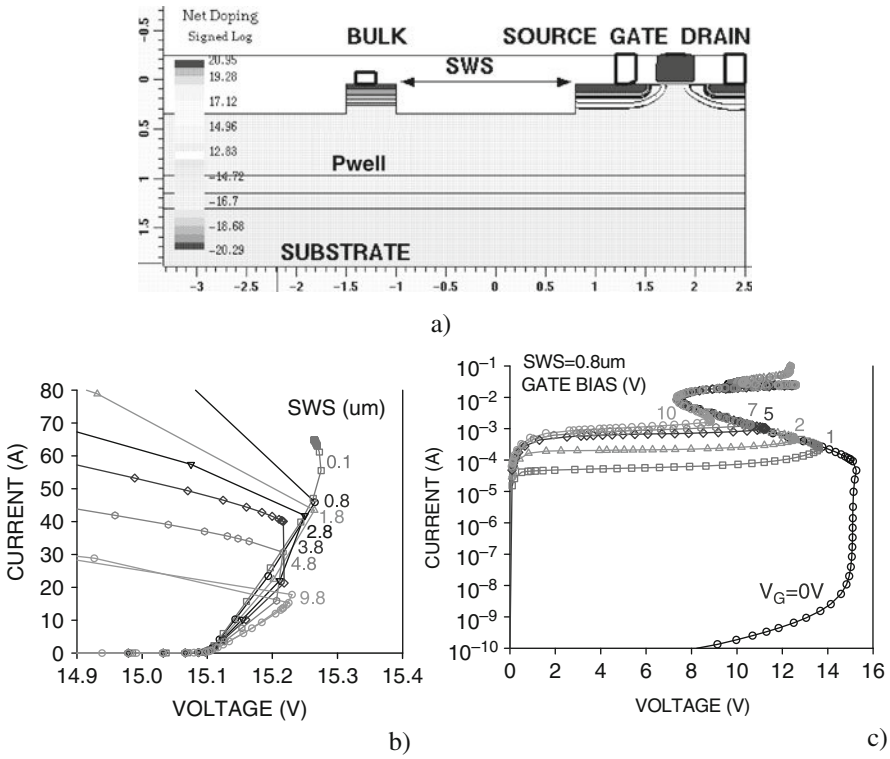


Fig. 7.21 Two-dimensional cross section for a finite element NMOS device used in mixed-mode TCAD analysis of a power array (**a**) and initially calculated quasistatic isothermal I_D–V_{DS} characteristics for different spacings between the P-well and source contacts (*SWS*) (**b**) and for different gate bias V_{GS} (**c**)

The original 5 V circuit demonstrated irreversible failure during switching after the elevation of the input power supply voltage VDD up to 6–6.5 V.

Analysis of the circuit uses the following assumptions for the particular circuit. The load resistor R_L is ~3.6 Ω for VDD = 5.5 V, and a duty cycle of 50% limits the output load current at ~(5.5/2)/3.6 = 0.763 A. The waveforms provided by the driver (not shown) apply a controlling "pgate" signal to the high-side PMOS MP, causing it to be in the low current state when the "ngate" signal is also low. Conversely, the "ngate" signal is provided at the high level only when the "pgate"

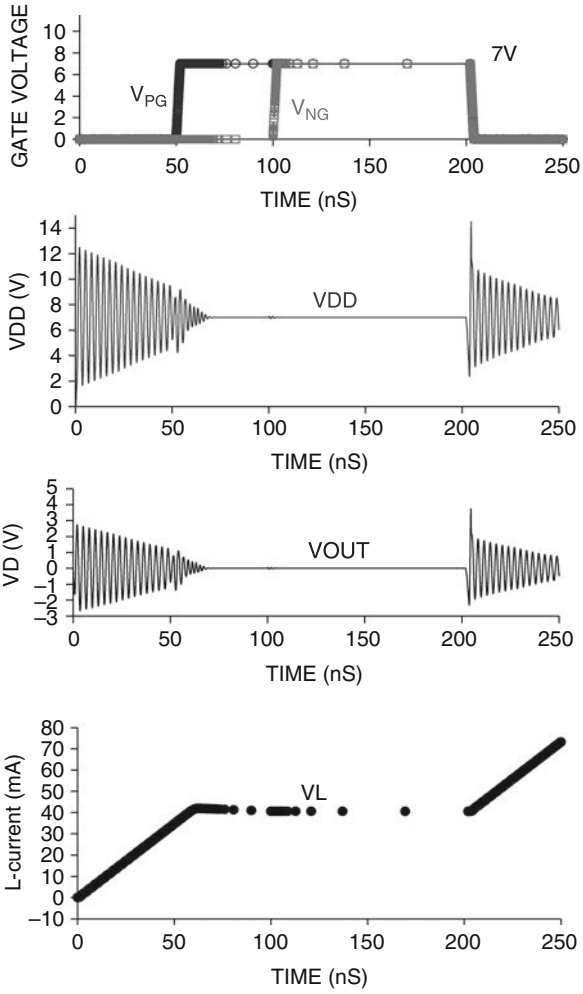


Fig. 7.22 Waveforms of the input gate voltages (*first graph*), VDD node (*second graph*), output node of the inverter (*third graph*), and the current through the inductance (*last graph*) for VDD = 7 V

signal is also high. The non-overlap time is about 50 ns (Fig. 7.20b). In the initial conditions, both the output NMOS and PMOS devices are in off-state. Then, the VDD node is set to some voltage level, followed by a preset of the output capacitor to the same voltage. This way, both terminals of the inductor are set to the same voltage as the output capacitor.

This major case study is related to the switching NMOS array layout effect. In the original design, the NMOS structure layout had random P-well contacts. The P-well diffusion to the source spacing followed the requirement for a 10 μm maximum space given by the process design rules. Such a design has been originally chosen as a space-saving measure, since it provides a high-density array placement.

7.2.3.2 Mixed-Mode Simulation

The mixed-mode numerical analysis has confirmed the experimentally observed failure scenario and helped validate the appropriate improvement measures. These measures involved the implementation of a fully butted NMOS array design with an additional redesign of array topology for appropriate current balancing.

The original 2D cross section for the finite element NMOS device used in the mixed-mode TCAD analysis of the power array is presented in Fig. 7.21a. The initially calculated quasistatic isothermal *I–V* characteristics for different spacing between the P-well contact (*SWS*) and for different gate bias (Fig. 7.21b) demonstrate consistent (Chapter 3) dependence of the critical voltage upon both the internal base resistance of the parasitic NPN structure and the channel current that provides the current in the drain multiplication region.

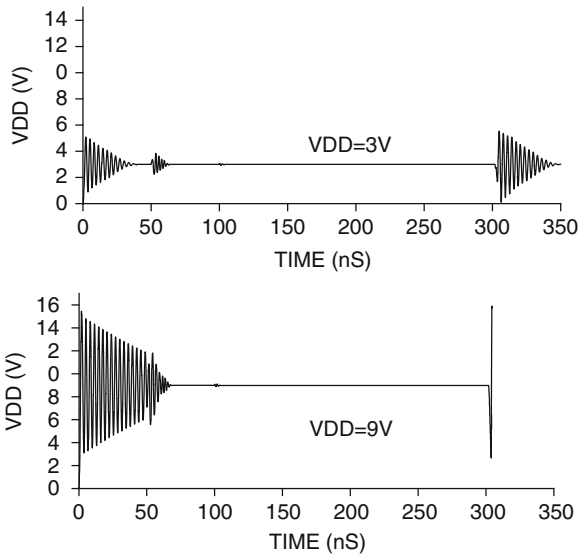


Fig. 7.23 Increase of the transient voltage noise with the input VDD voltage increase

As demonstrated by the simulation results, due to the inductance of the bond wires during circuit operation, the overstress of the NMOS array can significantly exceed the VDD level (Fig. 7.22). In this case, the amplitude of the transient voltage noise increases with the increase of the input voltage level (Fig. 7.23). This effect is assumed responsible for the early damage of the product between the operational and absolute maximum voltage range.

Using the mixed-mode simulation setup, it has been demonstrated that reduction of the p^+ to source diffusion spacing (SWS) to 0.1–0.8 μm results in an increase of the critical current density for snapback mode during switching by ~3–10 times

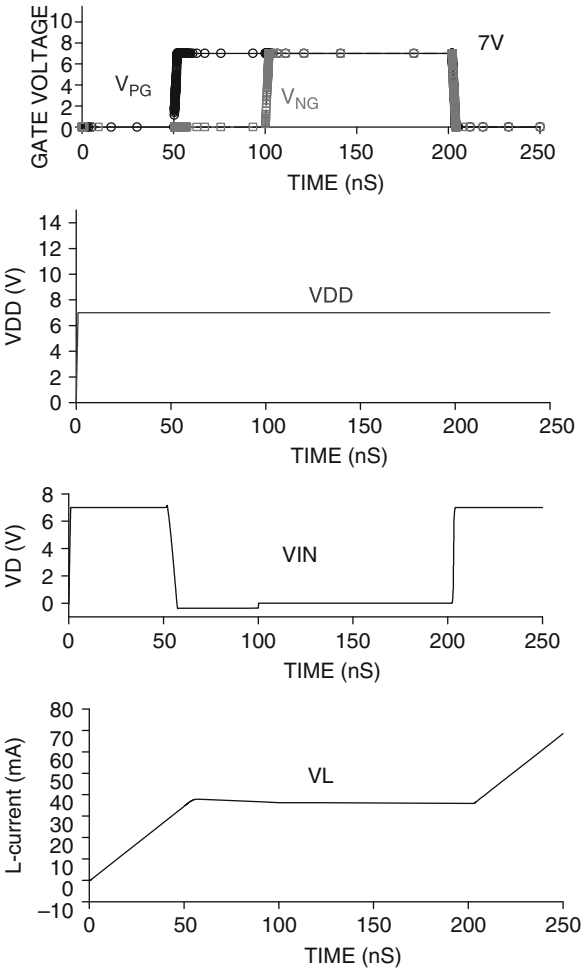


Fig. 7.24 Waveforms of the input gate voltages, VDD node, output node VL, and the current through the inductance for the case of eliminated bond wire inductance components

(or even more, considering an isolation effect by the intermediate fingers, which act similar to guard rings).

In addition, it is logical to assume that the 3D effects of the array's topology are important. The critical impact of the bond wire inductance components is demonstrated by the absence of transient voltage noise when bond wire inductances are removed (Fig. 7.24).

A similar conclusion can be reached using a compact model simulation with a snapback ESD compact model for the NMOS component. This approach allows for the inclusion of more driver components into the circuit.

7.3 ESD Protection of Integrated High-Voltage Regulators

7.3.1 Asynchronous Integrated Buck Regulator Case

7.3.1.1 Functionality and ESD Protection

A block diagram for the HV-integrated 100 V buck regulator is presented in Fig. 7.25. This asynchronous voltage regulator is based on a 100 V switching power NLDMOS array driven by internal circuit. The array supplies switching current pulses through the integrated LC filter to create a corresponding lower output voltage level (corresponding to the switching duty cycle).

The list of pins for package-level ESD protection includes the following control pins that can be assigned to the I/O group. The low-voltage 8 V pin R_{CL} is required to connect resistor for current limit control. The pins' absolute maximum voltage is 8 V and the 7 V snapback NMOS clamp provides an appropriate ESD protection solution.

The RTN pin is simply a power ground. The low-voltage VCC pin provides a regulated output at 7 V from the series pass regulator. This pin can be alternatively used to draw power from an external voltage source, which reduces the power dissipation in the circuit. Since the VCC pin is specified up to 14 V, it requires a 20 V snapback clamp.

The first high-voltage pin is the feedback pin. This pin receives a fraction of the voltage from the output divider. For a 100 V circuit, this voltage can be as high as 20 V. However, the voltage at feedback is already regulated, thus eliminating any fast transient conditions. Therefore, conventional middle voltage, 20 V-tolerant snapback protection presents an adequate solution.

For the high-voltage power group of the pins the requirements are the most challenging. The operational input voltage range VIN is from 9.5 to 95 V. Due to the high current during switching and the parasitic bond wire inductance, the voltage waveform at the input pad of the die might be significantly different from the package terminal, even if an external capacitor is used. The VIN pin requires high-voltage protection using a 100 V snapback clamp. Due to fast transient voltage, the clamp should have a very small dV/dt effect and use an avalanche diode reference. ESD protection of the pin can be achieved using snapback components

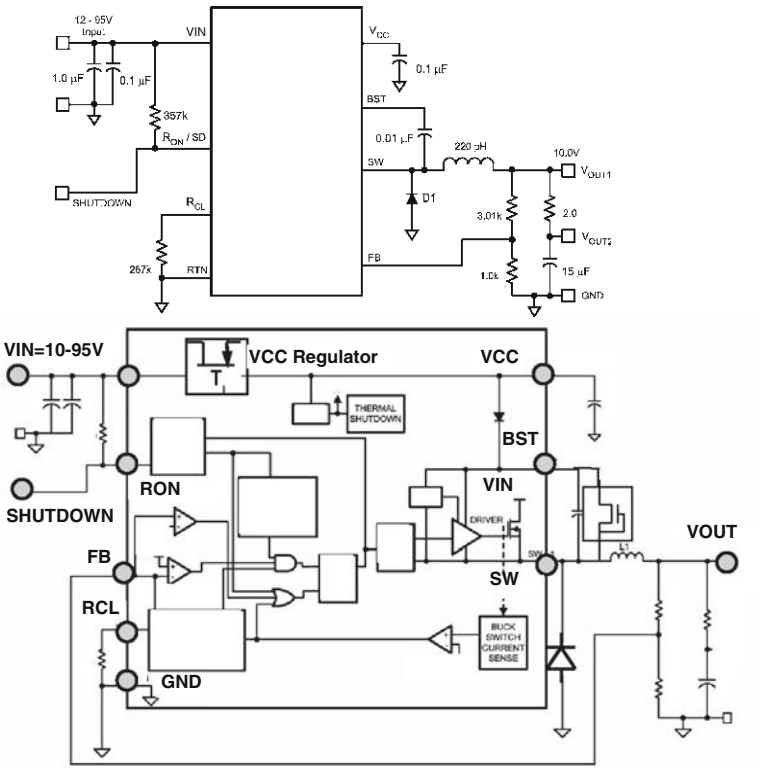


Fig. 7.25 Application circuit and functional block diagram of the 100 V buck dc–dc voltage regulator

with a holding voltage below the maximum input voltage (95 V). However, a more reliable approach involves the application of the high holding voltage lateral PNP ESD clamps (Chapter 4).

Protection of the switch pin SW is dependent on the capability of the array. In a 1 A power array with an appropriate design of the circuit layout, the switch pin can be self-protected by the power array. In this, most favorable, case, no ESD protection is required. An important consideration for the ESD current path is that the power MOSFET still contributes the body diode into the ESD current path, and thus appropriate metallization routing should be designed.

Finally, the boost-strap capacitor input (BST) is required to connect the output inductor, re-circulating diode, and bootstrap capacitor. The boost pin is an overvoltage pin. The voltage tolerance of the pin exceeds the switch node by the additional 7 V required for the adequate power supply for the driver. This is required to provide a typical boost functionality of on-state gate bias for the power MOSFET, including the conditions of low output voltage. If no higher voltage devices are available, then the only way to supply boost pin protection within the 100 V process limitations is to protect the pin relative to the switch pin. The corresponding level of high-voltage

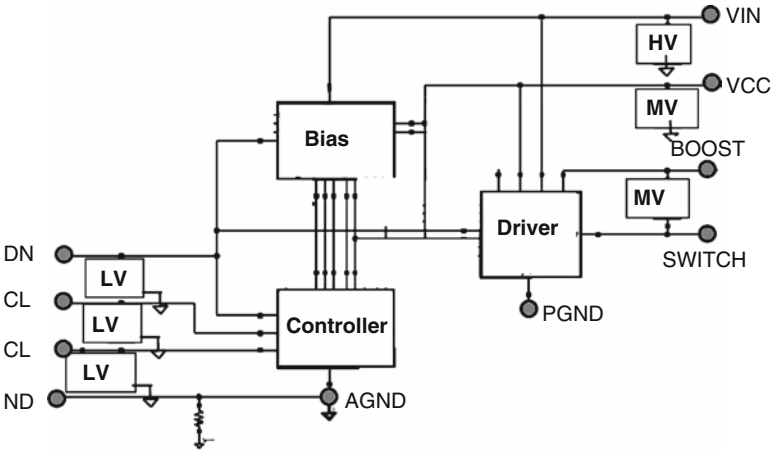


Fig. 7.26 ESD protection network for a 100 V asynchronous switching buck regulator

isolation that enables such network implementation is usually requested of process developers.

The final ESD protection diagram composed of low-voltage (7 V), middle-voltage (20 V), and high-voltage (100 V) clamps is shown in Fig. 7.26.

7.3.1.2 Case Study

For reference, the pulsed SOA area for the power-optimized 1 mm 100 V NLDMOS device is presented in Fig. 7.27a. In agreement with Chapter 4, the critical voltage of the device in pulsed conditions is a strong function of the gate bias. These characteristics can be measured for a small array using conventional TLP tools (Fig. 7.27b). However, the total current that array can withstand under ESD stress is a more nonlinear function of array topology and other factors, including parametric process yield.

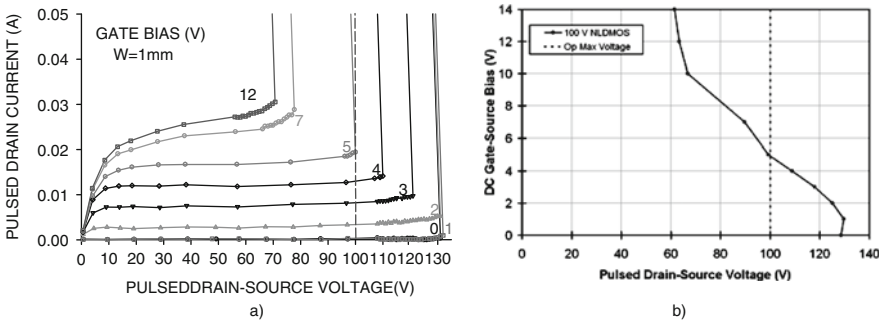
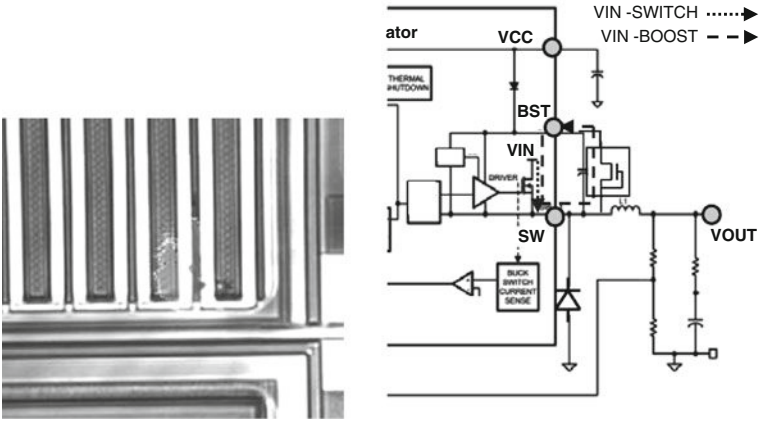


Fig. 7.27 Drain-source TLP characteristics measured at different constant gate bias (**a**) and plotted dependence of the gate bias upon the critical drain-source voltage (**b**) for a 100 V NLDMOS device

At the same time, the self-protection ESD current level of the array is a function of both the initial gate bias and the gate coupling level that can be tolerated by the gate driver.

Therefore, it is not a surprise that the original array design may not provide sufficient ESD robustness.

In particular, an example of machine model (MM) damage can be originally seen in the corner of the power array (Fig. 7.28a). This problem has been eliminated by redesign of the array. Two current paths were located by partition testing, revealing a VIN-SW and a VIN-BST current path (Fig. 7.28b). The redesign involved array layout optimization toward higher self-protection levels [136].



a) b)

Fig. 7.28 Original array showing damage after MM ESD stress (**a**) and ESD protection network for a 100 V asynchronous switching buck regulator (**b**)

Another challenge that typically causes practical problems is the protection of the internal VCC regulator (series 7 V regulator, Fig. 7.29).

The first current path formation leading to failure (Fig. 7.29a) involves HBM pulse coupling through a voltage divider that is composed of the parasitic drain–gate and gate–source capacitors on NLDMOS (Fig. 7.29b). In this case, the stacked NLDMOS devices are driven into a low-impedance state prior to HV ESD clamp activation.

In case of the second failure, the failure mechanisms have been initiated by the insufficiently high voltage of PMOS breakdown or leakage, leading to gate bias that turns on the NLDMOS prior to HV ESD clamp activation (Fig. 7.29a). This effect has been confirmed by electroluminescence of the PMOS device region using the PHEMOS technique (Fig. 7.29c).

The reliability improvement measures for these types of failure involve redesign of the VCC regulator using stacked devices, process-level fixes of the high-voltage PMOS, and implementation of more precise triggering characteristics for the high-voltage clamp.

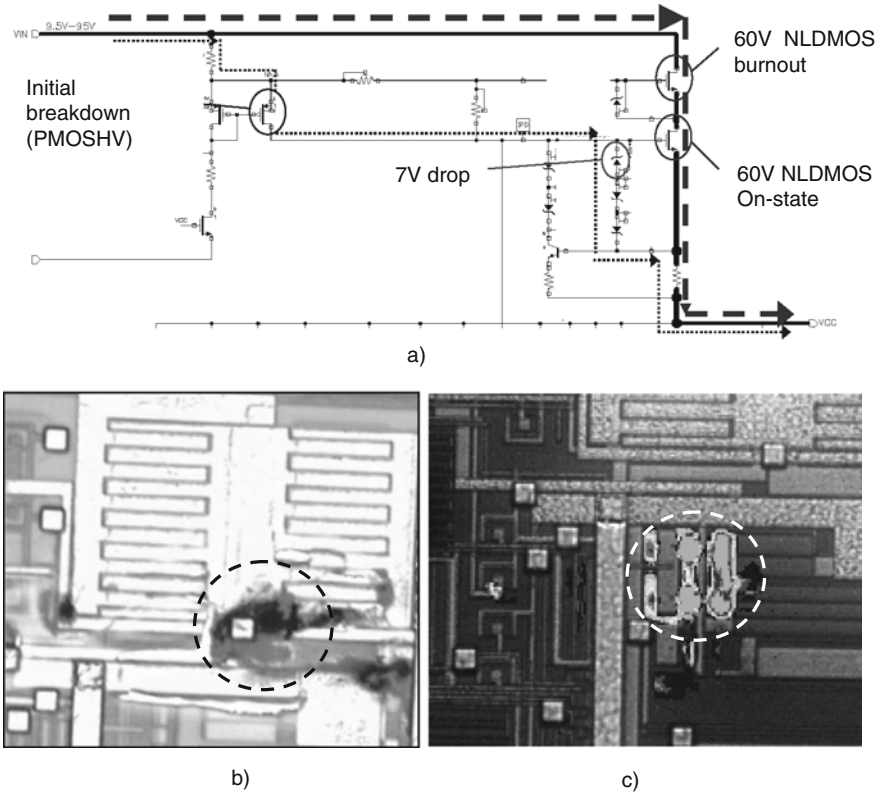


Fig. 7.29 VIN to VCC HBM failure current path (**a**), FA photo for NLDMOS failure under all pin HBM testing (**b**), and photoemission from PMOSHV at 110 V between VIN and GND pins (**c**)

7.3.2 Synchronous Regulators

7.3.2.1 HV Power Train Block

In a synchronous buck regulator, two high-voltage devices are used to realize the transient current path during the switching cycle. Instead of a diode to provide the current path through the LC filter, both the high-side power HV PLDMOS and the low-side HV NLDMOS devices are used in a synchronized turn-on and turn-off sequence with appropriate delay times (Section 7.1). The rest of the operation principle is similar to an asynchronous regulator and presents itself as an integration of the switching pulse with the input voltage amplitude VIN on the external LC filter in order to obtain the output voltage VOUT, regulated by changing on the switching duty cycle.

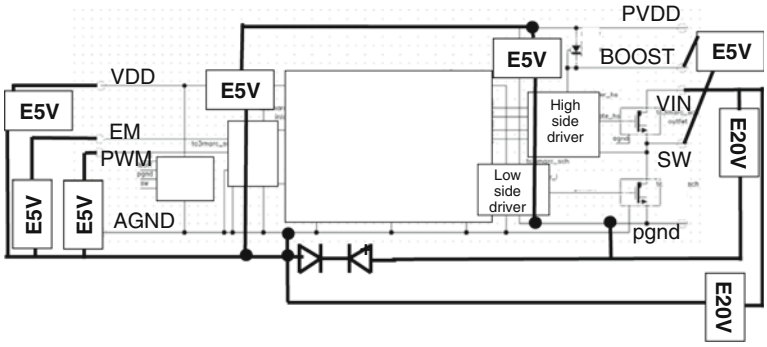


Fig. 7.30 A 24 V power train with ESD protection

A simplified circuit diagram and local clamp-based ESD protection network for a 24 V synchronous power train is presented in Fig. 7.30. The output low-side NLDMOS and high-side PLDMOS are driven by the 5 V low-side and high-side drivers, respectively. The low-side driver is powered by the PVDD external voltage. The supply voltage of the high-side driver is obtained from the voltage formed on the BOOST pin. The two domains, power and analog, have separate power (PGND) and analog (AGND) grounds that are separated for noise isolation by back-to-back diode clamps. The analog domain is powered by the VDD power supply.

The analog circuit block is represented by the corresponding power supply VDD pin, enable (EN) control pins, and pulse width modulation pin PWM.

Implementation of local ESD protection for this simplified power train circuit schematic practically represents ESD solutions for a wide variety of DC–DC products (Fig. 7.30). ESD protection of the low-side drivers is ground referenced and realized using a low-voltage snapback solution.

The high-side driver BOOST pin is protected relatively to the SW pin by treating it as a high-side ground. For this, isolated low-voltage snapback clamps are used. The current path is completed using HV local clamps between the PVIN and PGND and the PVIN and AGND with an optional clamp at the switch pin (Fig. 7.30). The HV clamp at the switch pin faces the same challenges and limitations as already described for LV power train local snapback protection.

Since control pins usually do not require low-impedance current path, two-stage protection is used for the input gate of MOS devices relying on current dissipation in the internal circuit with an additional current limiting polyresistor of ~1 kΩ. Alternatively, a second-stage 7 V avalanche breakdown diode can be used, which provides the corresponding current of ~1–10 mA to dissipate the second-stage current, thus dropping the pin voltage from the voltage level produced by the snapback NMOS clamp (~10 V) down to a safe level.

The power ground PGND is decoupled from the analog ground AGND, allowing it to be tolerant of relatively high voltage swings. This is accomplished by a back-to-back diode clamp.

The high-voltage switch (SW) and BOOST pins can be protected by a different high-voltage 24 V snapback clamp. Since high-voltage clamps are isolated from the P-substrate in a BCD process, the power input PVIN could be protected not only by a high-voltage clamp to the power ground PGND, but by a similar clamp to the analog ground AGND as well. Thus, Nepi isolation between the grounds is automatically provided.

In addition to power array protection, the most critical circuit block is the drivers where a match should be achieved between the 5 V snapback NMOS device waveforms and the midsize internal NMOS devices connected through the low-resistance current path.

Additionally to high-voltage node protection, the most challenging solution is required for the high-side driver. Since the boost pin follows the transient voltage at the switch pin, the low-voltage isolated clamp needs to have a low dV/dt effect. At the same time, power circuit requirements do not allow implementation of second-stage protection. As a result, targeting of the ESD protection window creates challenges.

In this case, the most typical failure is in the low-voltage NMOS and PMOS components circled in the fragment of the driver circuit in Fig. 7.31. Depending on the size of the driver components, two different current paths can be realized. In a high driver, M5: 25 μm $\times$ 12 = 300 μm and M6: 25 μm $\times$ 100 = 2500 μm. According to the simulation results with an ESD compact model (Fig. 7.32), the PMOS has a low gate bias potential and the current is directed into the power array gates. In the case of the low-side driver: M5: 50 μm $\times$ 20 = 1000 μm and M6: 50 μm $\times$ 261 = 13,050 μm, and the current is directed to the M5 and M6, resulting in damage of these devices.

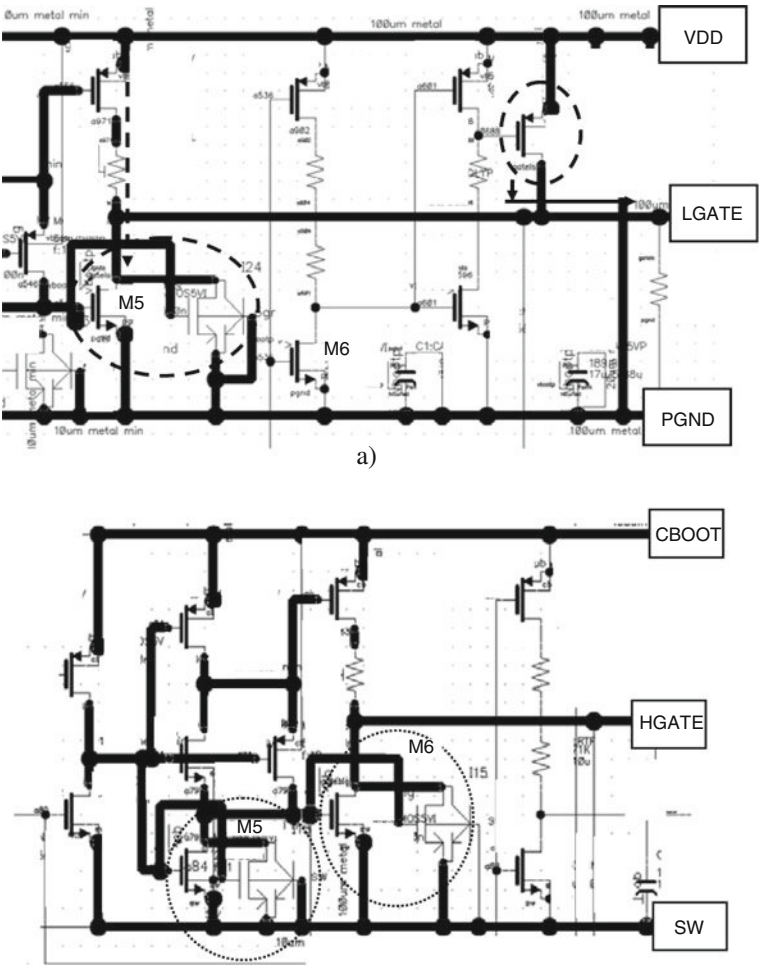
7.3.2.2 Synchronous Buck Regulator

The same principles used for power train ESD design are applied to the more complicated analog circuit of the synchronous buck switcher.

An example of a synchronous switcher that can accept an input voltage range from 3 to 20 V and provide up to 4 A of output current is presented in Fig. 7.33. In this particular simple switcher design, core logic is not in use and thus ESD protection of the low-voltage (2.5 V) pins is not required.

The circuit can be implemented in the non-copper top and copper top versions. In the case of copper top variety, all circuitry can be put under copper pads, including the ESD.

The major circuit pins are similar to the power train example discussed above. The internal chip has separate analog and power grounds that bond to the same package pin to reduce the bond wire impact. The analog ground AGND supports the global ESD bus across the analog block, while the power ground is decoupled using either the back-to-back or the stacked back-to-back NMOS clamp. The switch pin can be optionally protected by a precision high-voltage clamp. The same clamp is used in the 20 V-tolerant input pin. CBOOT pin protection is accomplished



a)

b)

Fig. 7.31 Overstressed devices in 24 V high- (**a**) and low-side (**b**) drivers

relative to switch pin using a low-voltage isolated snapback NMOS. The same protection is implemented for the VDD pin that supports the external connection to the decoupling capacitor.

The control high-voltage power good and enable pins are also protected by high-voltage clamps.

The feedback pin requires an elevated voltage tolerance of up to 10 V and thus uses a corresponding 10 V clamp.

The decoupling between the power and analog grounds is required since the current through the power block is significant.

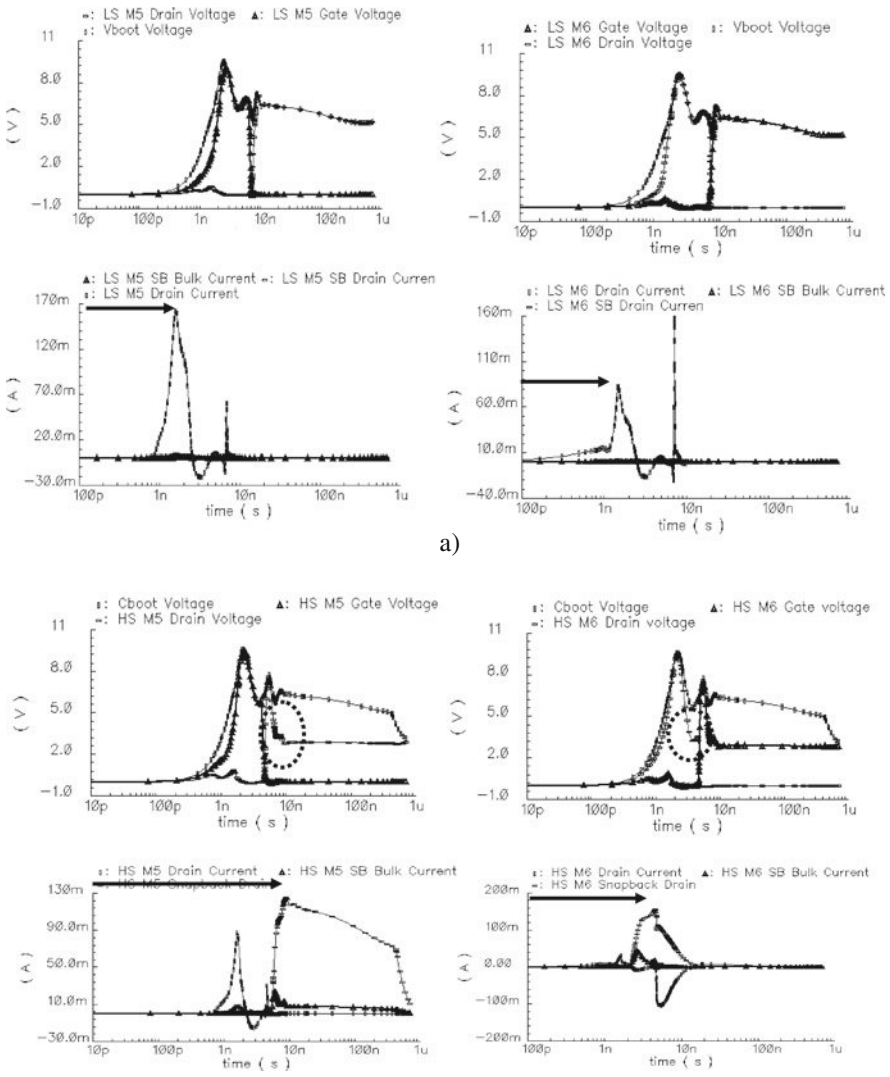


Fig. 7.32 Comparison of the 2 kV HBM simulation results for low-side (**a**) and high-side drivers (**b**)

The source of the low-side NLDMOS is connected to the PGND. It is, however, required that the PGND signal be able to swing up to ±5 V relative to the analog ground AGND. Respectively, the ESD test program includes an ESD zap combination of PGND to AGND. The possible current path is formed through the reverse-path diode or parasitic diode of the 5 V snapback NMOS clamp. In the case

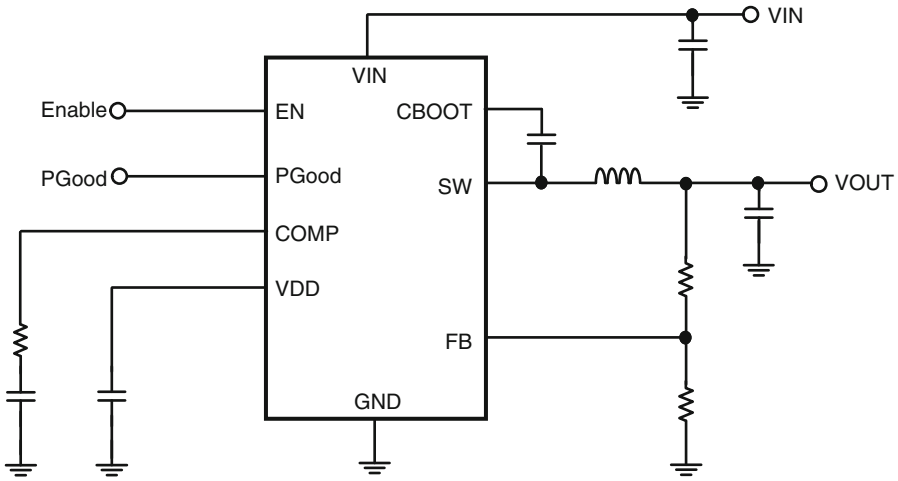


Fig. 7.33 Application for a simple switcher

of the opposite AGND to PGND zap combination, the reverse-path diode of the isolated NMOS snapback clamp conducts the current.

Protection of the input voltage pin is illustrated in Fig. 7.34. In the case of the input to VBOOT (similar to VDD) zap combination, the current path relies on the PLDMOS body diode. In the case of AVIN to analog ground AGND zap, the high-voltage snapback clamp is engaged (Fig. 7.34a).

Additional protection measures are required for the internal circuit, so 7 V avalanche diode clamps are extensively used across the internal circuit to clamp the gate–source voltages of the internal components (Fig. 7.34b), thus reducing the risk of a significant reduction of the pulsed critical voltage and gate oxide damage.

Protection of the power side is illustrated in Fig. 7.35.

The level shift circuitry is the most sensitive, as it is exposed to the full CBOOT voltage to AGND. To protect the internal voltage regulator, a separation resistor of ~10 kΩ is used (Fig. 7.36) or even an additional internal voltage clamp.

Another major concern is about the two high-voltage pins power good (PGOOD) and ENABLE. The challenge involves reliable protection of small, internal high-voltage devices interacting with the pin.

In this case, an adequate ESD protection window is needed to provide survivability of protected devices with a total width in the range of 100–300 μm. If the process technology provides an option of higher voltage components not optimized for the switching operation, such components will be the best choice. Another alternative includes circuit design based on stacked devices, in order to guarantee that the total voltage limited by the ESD protection clamp will not exceed the pulse SOA limits in the rather unpredictable conditions of gate coupling.

If none of this is possible, a separation polyresistor is the best choice (Fig. 7.37).

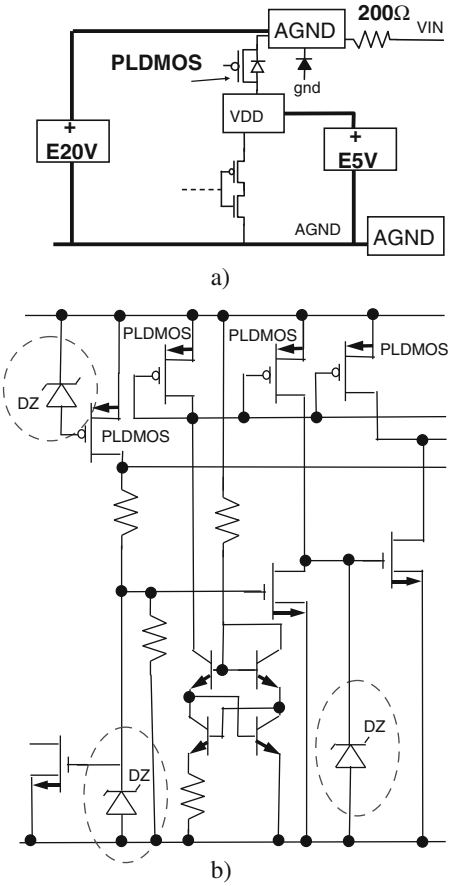


Fig. 7.34 Input voltage pin protection: AVIN to VDD and AVIN to AGND

7.4 Controllers

Controllers are integrated components that provide driver signals to the discrete switching components in hybrid dc–dc voltage regulators. Different types of dc–dc converters can be realized with the appropriate selection of the controller type, both the asynchronous and synchronous, buck and boost varieties.

In principle, ESD protection of the controller pins is rather similar to that of the integrated voltage regulator. The difference consists in the necessity of switch pin protection for synchronous converters. The driver (DR) pin is connected to an external discrete switch device. Often, output devices in the controller do not have sufficient size to realize the self-protection capability. Therefore, an explicit power clamp at the switch pin may be required. Under such conditions, it is important to know what is the gate coupling of the array in ESD stress conditions and what is the

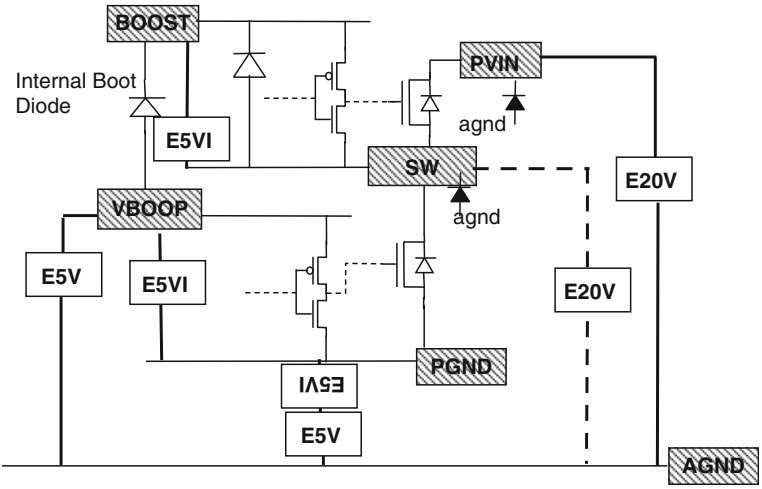


Fig. 7.35 ESD protection for the power-side SW with optional HV clamp

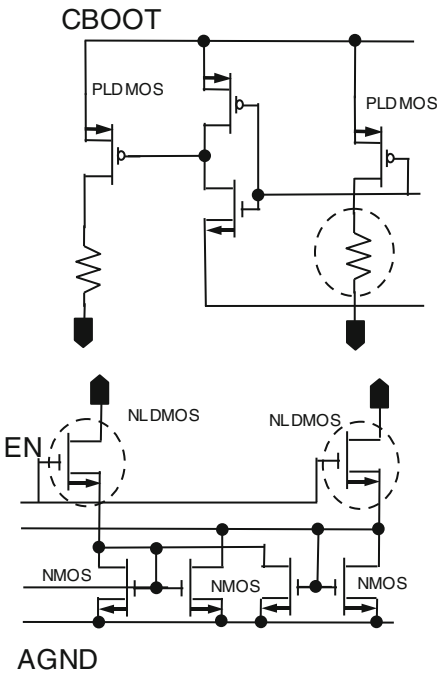


Fig. 7.36 Level shifter circuitry

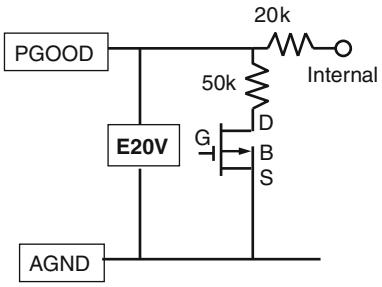


Fig. 7.37 Power good (PGOOD) high-voltage pin and ENABLE pin protection

pulsed SOA of the device. Several aspects of controller protection are highlighted below.

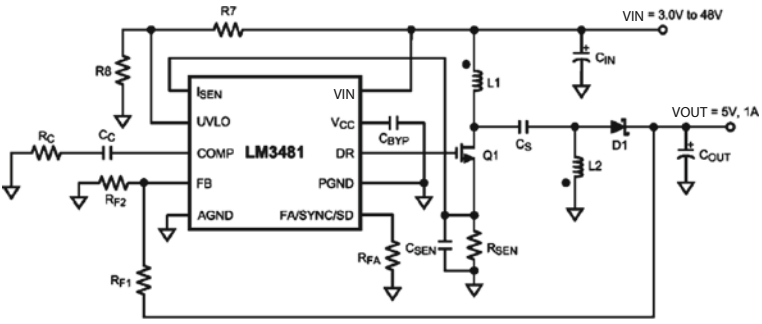
7.4.1 Asynchronous Buck-Boost (SEPIC) Controller

An example of the application and the functional diagram of a wide supply voltage range (2.97–48 V) controller are presented in Fig. 7.38 [126]. Typical applications of this controller include various distributed power systems, notebooks, PDAs, digital cameras and other portable applications, offline power supplies, and set-top boxes.

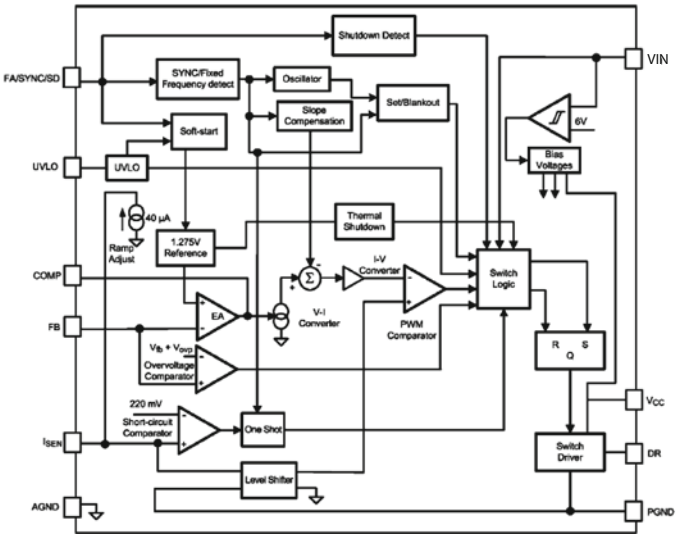
The controller drives the low-side discrete N-FET in switching regulators. It is suitable for use in topologies requiring a low-side FET, such as a boost, flyback, SEPIC. The controller can be operated at extremely high switching frequencies (up to 1 MHz) in order to reduce the overall solution size. The switching frequency is adjusted by using a single external resistor or by synchronizing it to an external clock.

Similarly, the current limit can be programmed with a single external resistor. Other capabilities of the controller include standard protection features such as thermal shutdown, short-circuit protection, and overvoltage protection. Another typical feature is a power-saving shutdown mode that reduces the total supply current to 5 μA.

The controller uses a fixed frequency with pulse width modulation (PWM) and current mode control architecture. In a typical application circuit, the peak current through the external MOSFET is sensed through an external sense resistor. The voltage across this resistor is fed into the ISEN pin. This voltage is then level shifted and fed into the positive input of the PWM comparator. The output voltage is also sensed through an external feedback resistor divider network and then fed into the error amplifier (EA) negative input (feedback pin, FB). The output of the error amplifier (COMP pin) is added to the slope compensation ramp and fed into the negative input of the PWM comparator. At the start of any switching cycle, the oscillator sets the RS latch using the SET/blank-out and switch logic blocks. This forces a high signal on the DR pin (gate of the external MOSFET) and the external MOSFET turns on.



a)



b)

Fig. 7.38 Application circuit (**a**) and functional block diagram (**b**) of the low-side NMOSFET controller

When the voltage on the positive input of the PWM comparator exceeds the voltage on the negative input, the RS latch is reset and the external MOSFET turns off.

The voltage sensed across the sense resistor generally contains spurious noise spikes. These spikes can force the PWM comparator to reset the RS latch prematurely. To prevent these spikes from resetting the latch, a blank-out circuit inside the IC prevents the PWM comparator from resetting the latch for a short duration after the latch is set. This duration, called the blank-out time, is typically 250 ns and is specified as t_{min} (on) in the electrical characteristics section. Under extremely light load or no load conditions, the energy delivered to the output capacitor when the external MOSFET is on during the blank-out time is more than the energy

delivered to the load. An overvoltage comparator inside the circuit prevents the output voltage from rising under these conditions by sensing the feedback (FB pin) voltage and resetting the RS latch. The latch remains in a reset state until the output voltage decays to the nominal value. Thus, the operating frequency decreases at light loads, resulting in excellent efficiency.

The specific pins for ESD protection of converters are somewhat similar to the integrated switching regulators. VIN is usually decoupled from the power FET devices and requires a high-voltage power clamp.

The control pins can in most cases tolerate an additional separation resistor in the range of 1–5 kΩ. Thus, the protection can be based on a two-stage approach that takes into account a certain current that can be dissipated by the internal circuit components connected to the control pin.

Similar to the integrated converters, the feedback pin usually experiences a rather slow transient voltage and can be easily protected even by the d*V*/d*t*-triggered clamp.

The overall ESD protection network for the controllers is shown in Fig. 7.39. The ESD protection network is composed of one 50 V high-voltage power clamp at the VIN pin and a low-voltage clamp referred to the ground of the corresponding (power or analog) domain.

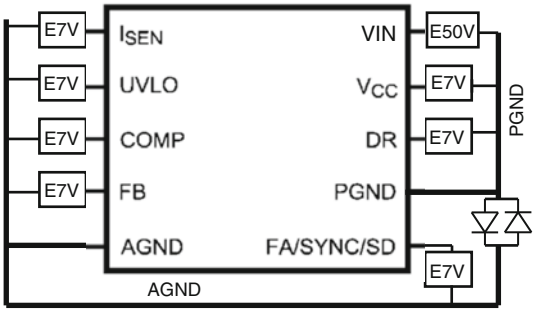


Fig. 7.39 ESD protection network for an asynchronous controller

The low-voltage control pins of the analog domain include a current sensing pin to detect the voltage generated across the external resistor, under voltage lockout pin (UVLO), the compensation pin (COMP), the feedback pin (FB) connected to the inverting input of the error amplifier, and the frequency adjust, synchronization, and shutdown pins. All of these pins have local protection with a separation resistor relative to the analog ground node.

The low-voltage pins in the power domain include the driver output (DR) and driver supply voltage pins used for decoupling the capacitor connection (VCC).

Thus, the overall ESD protection solution is based on a classical local clamp approach (Chapter 5). The most challenging pin is the high-voltage pin VIN. In spite of a much lower transient voltage in the controller (in comparison with the integrated voltage regulator), this pin still requires precise targeting of the clamp's turn-on voltage if the snapback solution is used. The targeting of precise ESD clamp

triggering is challenging in case of insufficient marginal SOA of the high-voltage devices supported by the process. Under these circumstances, specifically device level, ESD design is critical for the success of the protection in the final analog circuit product.

7.4.2 Synchronous Buck Controller

An example of the application diagram for a synchronous buck controller is shown in Fig. 7.40. This controller is intended for step-down regulator applications from a high voltage or widely varying input supply. The control method is based on current mode control utilizing an emulated current ramp. Current mode control provides inherent line feed-forward, cycle-by-cycle current limitation, and ease of loop compensation. The use of an emulated control ramp reduces noise sensitivity of the pulse width modulation circuit, allowing reliable control of the very small duty cycles necessary in high input voltage applications.

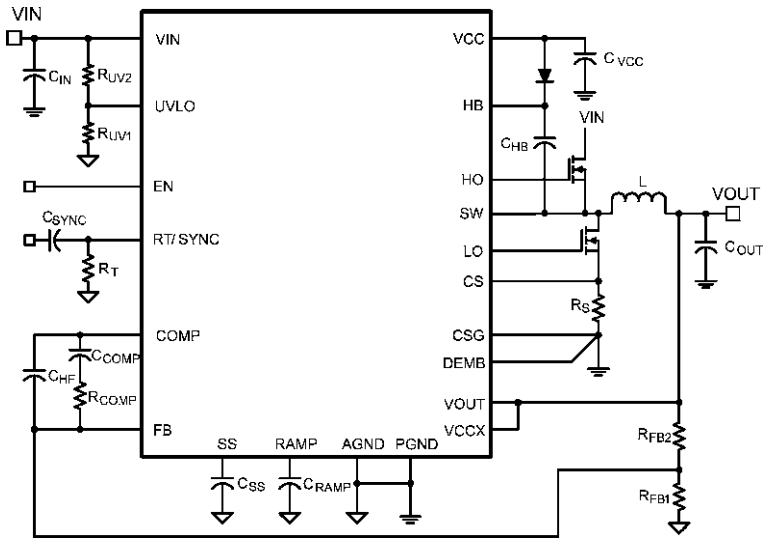


Fig. 7.40 Application diagram for 100 V synchronous controller

The operating frequency is programmable from 50 kHz to 1 MHz. The controller drives external high-side and low-side NMOS power switches with adaptive dead-time control. A user-selectable diode emulation mode enables discontinuous mode operation for improved efficiency at light load conditions. A low quiescent current shutdown disables the controller and consumes less than 10 μA of the total input current. Additional features include a high-voltage-bias regulator, automatic switchover to external bias for improved efficiency, thermal shutdown, frequency synchronization, cycle-by-cycle current limiting, and adjustable line under-voltage lockout.

The ESD protection network for the controller can be constructed based on the absolute maximum voltage limits for the circuit pins and the internal circuit blocks, using local 7, 16, and 100 V clamps for ESD protection. The network, with corresponding clamps E7V, E16V, and E100V, is shown in Fig. 7.41.

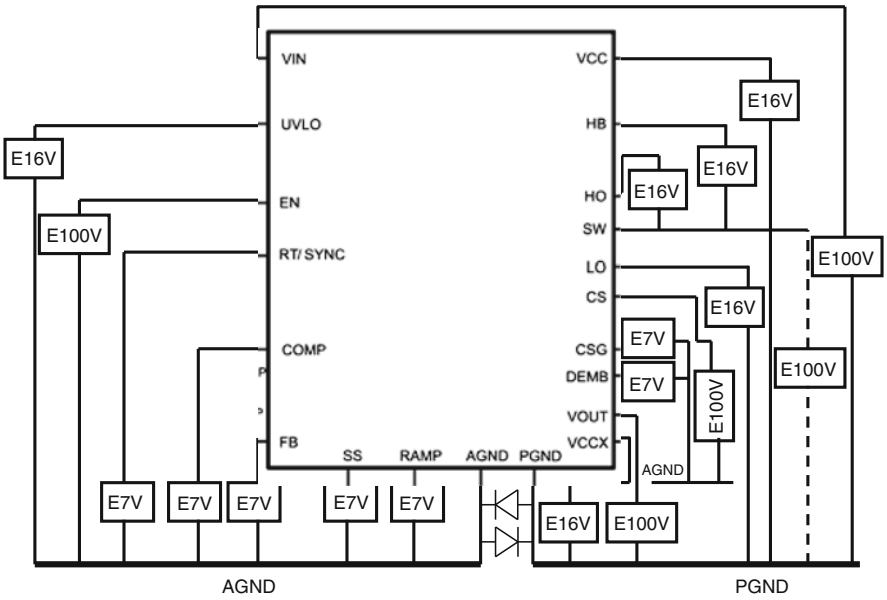


Fig. 7.41 ESD protection network for a 100 V synchronous controller composed of 7, 16, and 100 V local clamps

Even though both power and analog grounds are connected to the same PCB node in the application, decoupling of the analog and power domains is required inside the chip due to bond wire resistance and inductance. Correspondingly, the domains are separated by the back-to-back ESD diodes. The remaining network structure is based on local clamp protection, taking into account that each clamp incorporates the reverse diodes.

There are three reference-ground-like levels by which local ESD protection is organized. In the power domain, the VIN pin is functionally protected by the 100 V power clamp relative to the power ground. This pin also connects the chip voltage supply, which also connects the input voltage monitor and the input of the VCC regulator. Therefore, all the devices connected to the pin should be evaluated for alternative current paths.

Similarly, 100 V local clamp protection is applied to the VOUT and, optionally, to SWITCH pin.

Other pins with 100 V absolute maximum ratings are the current sense (CS) and enable (EN) pins, which use local 100 V protection relative to the analog ground. These pins can usually tolerate a second-stage resistor.

The power-side 16 V pins include an optional input for the external power supply VCCX, a VCC pin for local decoupling by the external capacitor, low- and high-side driver outputs LO and HO, and a boost pin (HB). The low-side driver output LO is protected using a power-ground-referenced 16 V clamp. The high-side driver and boost are protected relative to the switch, treating this fast transient pin as a local ground.

The other analog domain control pins are protected locally, relative to the analog ground (Fig. 7.41).

One of the peculiarities of some synchronous controller circuit designs is the requirement that a high-side switch pin must go below the power ground. In this case, it is important to avoid parasitic injection of the carriers into the substrate. An example of the solution is shown in Fig. 7.42. It provides for SW pin biasing below ground at a low current and voltage of up to 3 V. This dual-direction protection can be achieved by stacking high-voltage clamps that incorporate an additional base–emitter diode provided by the NPN BJT.

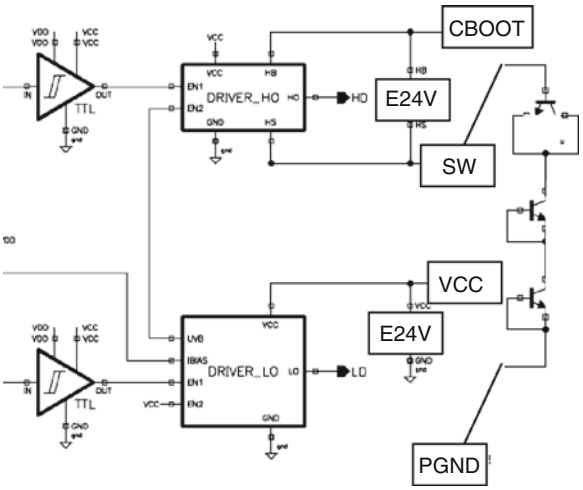


Fig. 7.42 Simplified block diagram for a 50 V controller circuit with ESD protection components

7.5 Light Management Units and LED Drivers

7.5.1 Analog LED Technology

Today, backlight light emission diode (LED) technology is replacing other light sources for liquid crystal displays (LCD). RGB (red, green, and blue) LED improves the color saturation on TVs and allows dynamic dimming. LED backlight technology has attracted a great amount of attention due to its wide color range, easy

application of dynamic driving, and scalability to larger sizes. Over time, cost-efficiency issues are being resolved by successful commercialization, for example, in LCD TV applications.

There are numerous applications for LED backlighting. For example, automotive cockpit displays, rear seat entertainment systems, computer displays; mid-size LCD TVs, digital frames; portable ultrasound, touch pads, industrial operator interfaces, security monitoring and video.

There are several ways to produce backlighting in systems that provide different spaces for LED drivers. Direct lit approach requires a small number of high-power LED modules with a current of up to 5 A. This approach can be realized using a large number of LED modules with small currents of 50 mA. Such an architecture provides advantages in thermal and optical aspects, but makes control more complex. The side-emitting approach requires a high brightness and a current of ~0.5 A when operating at some medium number of LED.

From an ESD point of view, the major concern created by the LED driver specifications is the current level of the output current sink components. In the case of the 50 mA output, each high-voltage current sink will require individual ESD protection, while in the case of single-point direct-lit drivers, the current sink power array might be already self-protecting.

The real light management units may be rather complex systems that include several modules (Fig. 7.43). In addition to the dedicated LED light source, such a system can include a multichannel color sensor, a temperature sensor, sophisticated digital modules for backlight control, and integrated power supply modules.

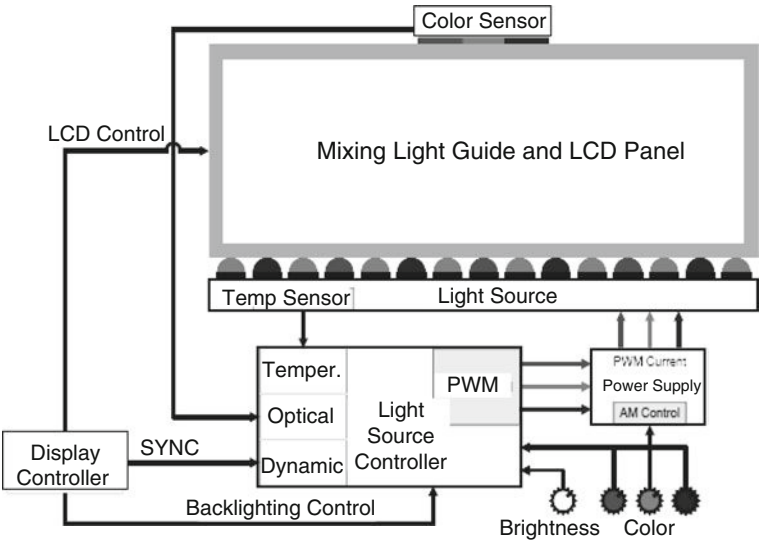


Fig. 7.43 Simplified example of the closed-loop basic architecture for LED backlighting

7.5.2 LED Drivers

An example of the application and functional block diagram for a constant current boost LED driver with internal compensation is presented in Fig. 7.44.

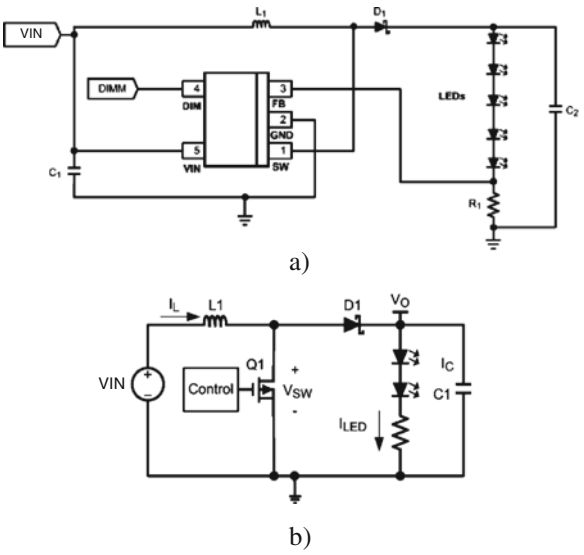


Fig. 7.44 Application (**a**) and functional (**b**) diagram of a LED driver circuit

The driver is intended for applications in the LED backlight current source, backlight OLED and HB LED drivers, handheld devices, LED flash drivers, and automotive uses.

The constant current LED driver is a monolithic, high-frequency, PWM DC/DC converter with a design that requires minimum external components. It can deliver typical 3 A peak currents with an internal 170 mΩ NMOS switch. The switching frequency is internally set to either 525 kHz or 1.60 MHz, allowing the use of extremely small surface-mount inductors and chip capacitors. Even though the operating frequency is high, efficiencies of up to 88% are easy to achieve. External shutdown is included, featuring an ultra-low standby current of 80 nA. The driver utilizes current mode control and internal compensation to provide high performance over a wide range of operating conditions. Additional features include dimming, cycle-by-cycle current limiting, and thermal shutdown.

The device operates similar to a voltage-regulated boost converter, except that it regulates the output current through the LEDs (Fig. 7.44b). The magnitude of the current is set with a series resistor.

The ESD protection network for this small pin count circuit includes low-voltage VIN protection relative to the power ground and protection of the control and feedback pins relative to the analog ground. Due to a very high current of the output NLDMOS array (over 2 A) the output (SW) pin does not require explicit protection. The ESD protection network described above is shown in Fig. 7.45.

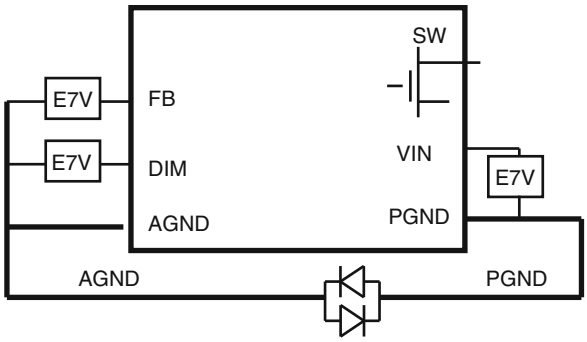


Fig. 7.45 ESD protection network for LED driver

7.5.3 Light Management Units

Unlike powerful LED drivers, the backlighting LMUs (light management units) are usually designed to control relatively low current levels through multiple LED strings. Respectively, the most specific pins for ESD protection are the high-voltage current sink pins. The LMU can be designed for both portable and integrated systems. In portable electronics, the most typical requirement for LED drivers is operation from a standard battery power supply of ∼3–5.5 V.

To provide high performance in backlighting, one of the most important parameters is the output voltage of the LED drivers. A higher voltage level provides a corresponding advantage in the stacking of a larger number of LEDs. This allows reduction of the number of wires in the portable system and simultaneously provides an equalized intensity of the light emission across the LED string, simply due to the same current level. Which in turn makes it possible to use the less-expensive LED diodes with unmatched electrical characteristics.

The LMU uses an integrated current mode pulse width modulation (PWM) boost converter, similar to the LED driver in portable applications. The converter becomes functional at an input voltage from 2.7 to 5.5 V and provides a regulated, high-voltage output to the LED string.

In addition to the voltage boost, the light management units (LMUs) can incorporate many features provided by the corresponding functional blocks, for example, programming via an I^2C-compatible interface.

A simplified circuit diagram for the LMU system is presented in Fig. 7.46. The driver includes an integrated boost voltage regulator to increase the battery input voltage level to ∼40 V, in order to produce power for a string of LED. Simplified internal circuit diagrams are shown in Fig. 7.47.

The module uses an external reference voltage source that is encoded by an analog–digital converter (ADC). In each channel, the LED intensity is controlled by the current sink circuits that use the pulse width modulation principle.

The LED bias supply circuit adaptively regulates the supply voltage of the LED strings to maximize overall efficiency. The biasing circuit provides regulated current

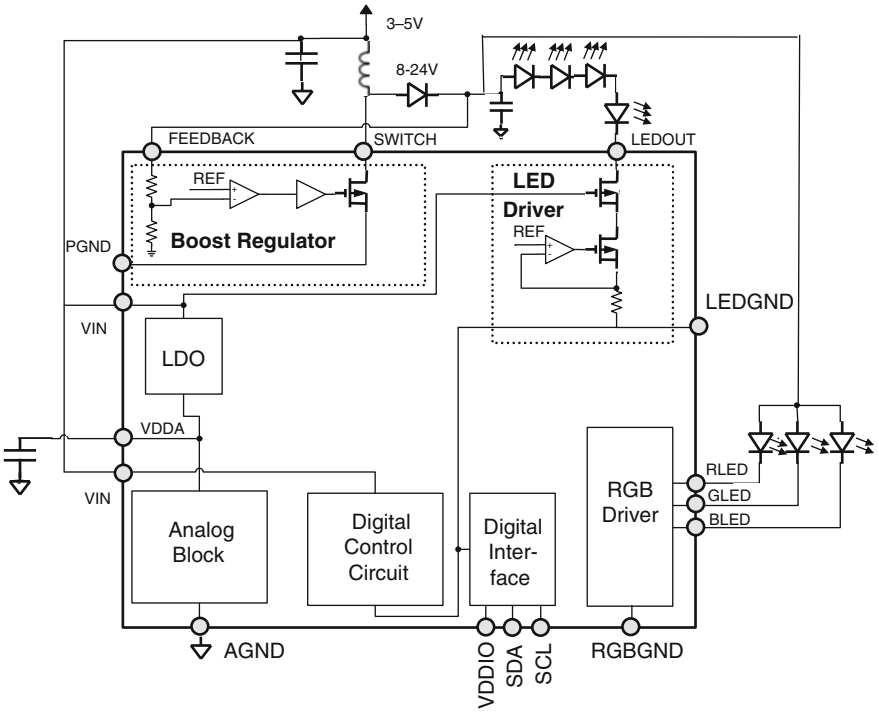
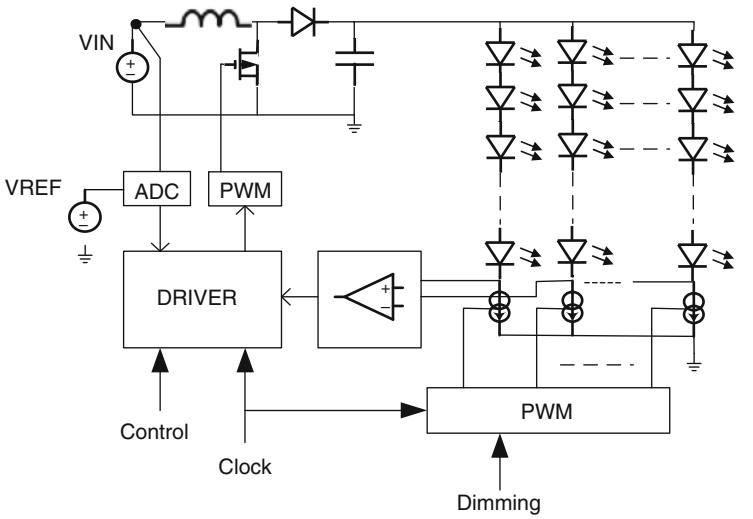


Fig. 7.46 Application circuits and functional block diagram for LMU

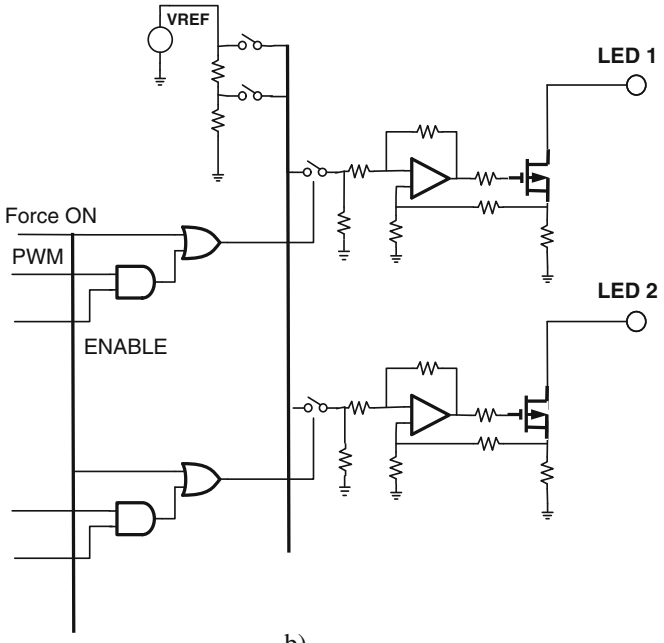
to the current sink circuit. The maximum current per output is set via an external low-power resistor connected to the control pins. An I^2C-compatible interface allows for independent adjustment of the LED current in either output.

The core of the circuit is a PWM current mode boost converter. At the start of each switching cycle, the internal oscillator sets the PWM converter. The converter turns the NMOS switch on, allowing current to ramp up in the inductor while the output capacitor supplies power to the LEDs. The error signal at the output of the error amplifier is compared to the sensed inductor current. When the sensed inductor current equals the error signal or when the maximum duty cycle is reached, the NMOS switch turns off, causing the external Schottky diode to pick up the inductor current. This allows the inductor current to ramp down, causing its stored energy to charge the output capacitor and supply power to the load. At the end of the clock period, the PWM controller is set again and the process repeats itself. When biasing LED strings, the circuit maximizes efficiency by adaptively regulating the output voltage.

The main ESD-specific requirement of LMUs is the protection of current sink pins (Fig. 7.47). For portable applications, the current sink is usually designed to supply a relatively low current of up to 30 mA to the LED stack. Thus, an open-drain current sink device requires special measures for ESD protection.



a)



b)

Fig. 7.47 Simplified circuit diagrams for LED backlighting block (**a**) and current sink circuits (**b**)

7.5.3.1 Switch Pin Protection

If the LMU is designed to supply current to multiple LED strings, the boost switch has a relatively large NLDMOS device. If the size of the output array is insufficient for self-protection, then local switch-pin protection is used. In the case of a very limited ESD protection window, two high-voltage clamps can provide an appropriate voltage limiting relative to both the power and analog grounds (Fig. 7.48).

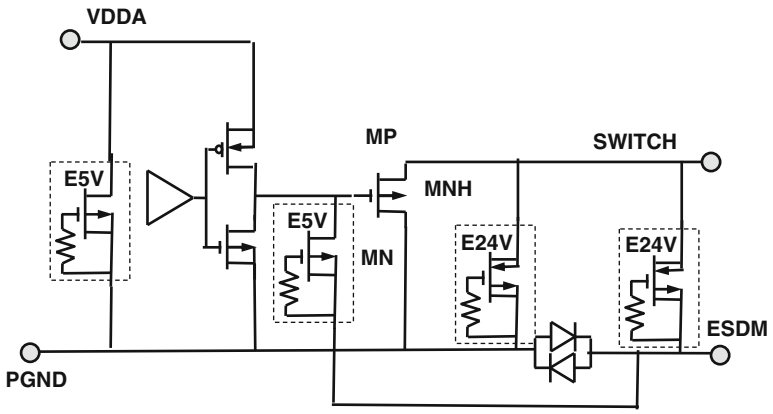


Fig. 7.48 Local ESD protection of the boost switch circuit

At the same time, since the array can transmit the ESD signal due to a high gate–drain voltage/current, additional 5 V snapback or avalanche diode protection can be used to clamp the switch gate of the NLDMOS and thus protect the driver circuit components. Respectively, the driver's output NMOS and PMOS transistors have butted well ties to maximize the critical avalanche current in the whole gate bias range (Fig. 7.48).

The switch pin experiences a very fast voltage transient when the switch is turned off. Non-interference of the ESD device with normal operation is crucial for the final analog product. The high-voltage clamp should be carefully designed or selected with features that eliminate the possible dV/dt effect and guarantee the absence of false triggering.

7.5.3.2 Feedback Pin Protection

Another fast transient circuit is the boost feedback circuit. It consists of a current sink and a feedback resistor divider. The sink current is only approximately 5 mA. This high-voltage circuit block requires a corresponding high-voltage local clamp (Fig. 7.49).

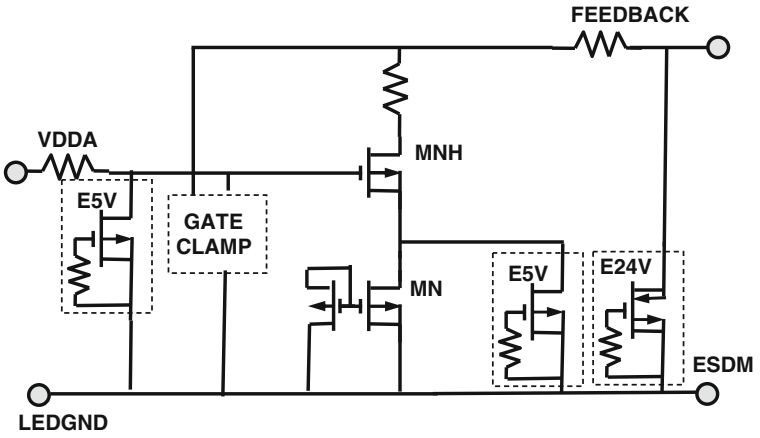


Fig. 7.49 Simplified circuit diagram for boost feedback circuit with ESD protection

7.5.3.3 LED Driver Protection

The LED driver (Fig. 7.50) is the most specific pin for LMUs. Normally, LED driver protection would require just a high-voltage clamp. However, in some cases, the power-optimizing process may provide no other alternative but to implement a cascade NLDMOS driver (Fig. 7.50). This is one of the direct examples of ESD and analog circuit co-design. Essentially, the only reason for stacked driver design is to enable protection of the low side NLDMOS driver under ESD stress. This circuit solution simply increases the absolute maximum voltage of the circuit, thus increasing the ESD protection window.

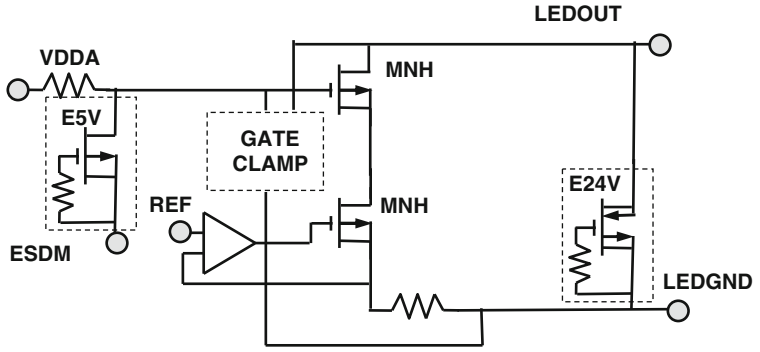


Fig. 7.50 Simplified circuit diagram for LED driver with ESD protection

Additional improvement toward robust protection includes the gate clamp described below and internal driver protection.

7.5.3.4 Gate Clamp

Measures for current sink and switch protection can be insufficient due to excessive coupling of the array. To resolve this issue, the capacitive coupled gate clamp is presented in Fig. 7.51; it is based on low-voltage NMOS devices MN1-MN3. During ESD stress of the boost pin, the capacitive coupling provides on-state conditions for the NMOS MN3 connected to the PGATE node, thus creating short-circuit conditions for the power NLDMOS (not shown on Fig. 7.51) in the boost circuit. As a result, the NLDMOS remains in the off-state during the stress, providing the highest possible drain–source triggering voltage. In this case the ESD current path is provided by the additional explicit high-voltage power clamp (not shown on Fig. 7.51). The CMOS diode of NMOS MN2 is used in the circuit to ensure that the excessive coupling through the capacitor will not damage the NMOS gate.

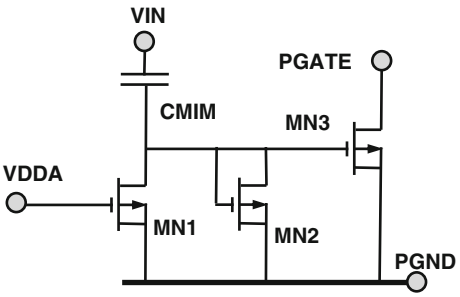


Fig. 7.51 Gate clamp

During normal powering sequence, in spite of the voltage change at the boost pin, the coupling capacitor remains fully charged, while the VDDA signal keeps the NMOS MN1 connected to the GATE node in the off-state.

7.5.3.5 RGB Driver

The RGB LED driver design includes cascaded NLDMOS (MNH) and low side low voltage NMOS (MN) connected to the driver (Fig. 7.52). In this case, a low-voltage NMOS is used due to the poor matching of the NLDMOS devices. In addition to 24 V protection with snapback clamp (E24V) the 5 V snapback clamp (E5V) is employed to ensure that the low-side low-voltage NMOS (MN) is protected (Fig. 7.52).

A resistor is needed for the cascaded NLDMOS (MNH) to ensure the functionality of the RGB driver (Fig. 7.52). This gate biasing is common for all RGB outputs. A 5 V clamp added to the NLDMOS (MNH) gate, as well as a capacitive coupled gate clamp. The capacitor is a capacitive connection from the output, with metal-to-metal capacitance if required, of approximately 150 fF.

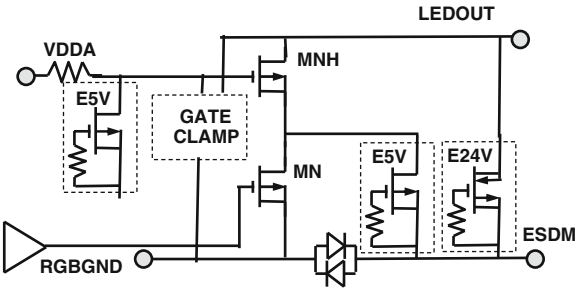


Fig. 7.52 Simplified circuit diagram for the RGB driver with ESD protection

7.5.3.6 Control Pins

The approach for control pin protection in LMU is similar to that of other analog circuits. Each specific pin design requires an analysis of the connected control circuit. In most cases, the control pins include a CMOS input buffer to provide the input function and an open-drain NMOS device to provide the output function.

For the output driver, usually even a relatively small (50 Ω) well saturation resistor between the pin and the NMOS drain provides a significant advantage (Fig. 7.53).

The high-voltage control pins may have an open-drain current sink as an output and an NLDMOS protected digital input.

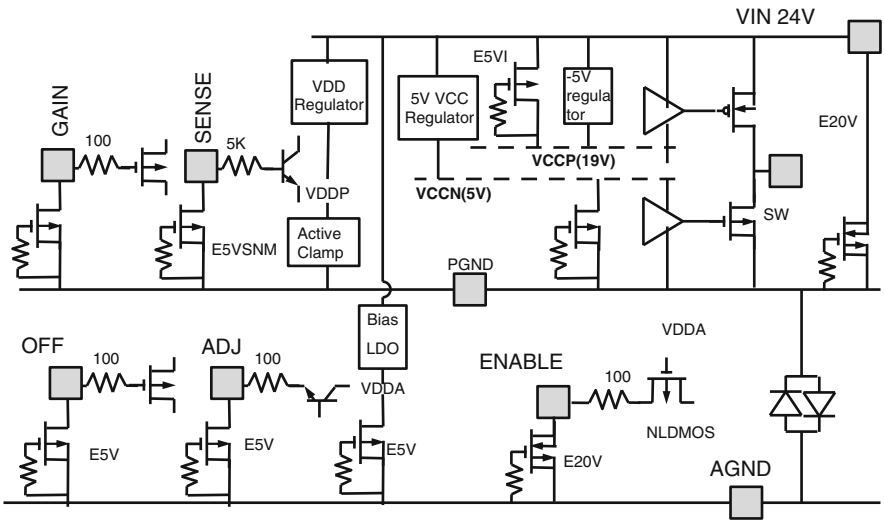


Fig. 7.53 Protection of the control pins

7.5.3.7 Current Sink Protection

The output of the main current sink is tied to the drain of a 2 mm NLDMOS device with ~4 Ω source degeneration resistor R_{ND}. Additional attention is required for isolation of the input devices so that they do not share composite. The shared composite of the two NMOS devices connected to two different pins can be problematic with respect to ESD due to paretic current path.

In the case of micro-SMD, the custom modifications of ESD cells include the removal of pad plates and the implementation of appropriate connections to the bump metal to avoid the current crowding effect.

Similar to boost regulators and controllers, protection of the boost pin can be done relative to the switch pin using an isolated low-voltage clamp. An additional alternative current path can be introduced by the overvoltage clamp between the boot and the power ground pin. Similar to the switch driver, additional measures have been used to increase the pulsed SOA of a 5 V NMOS device by butting the P-well ties to the source.

An alternative option is to implement the self-protecting capability of the LV NMOS by a drain extension of 1–3 μm. I/O latch-up prevention rules also apply, including double guarding.

7.6 A Few More Case Studies

In this last section of the chapter three extra case studies are discussed to emphasize one more time a non-trivial nature of the power management circuit design as well as to demonstrate the value of mixed-mode simulation approach. The cases cover quite unique events of transient latch-up and device failure under charged device model (CDM) pulse. All of them occur due to an unexpected current path in power analog circuit. The specifics of these cases are substantial dependence of the phenomena on IC layout and even on PCB design.

7.6.1 Power Array–ESD Clamp Interaction

In general the challenges for ESD protection of the power circuits are not limited by the ESD protection window targeting only. In principle the substrate current, generated by any internal circuit components, can impact critical triggering regime of the snapback ESD device both in ESD pulse domain and under other conditions. The scenario of the "accidental" ESD device turn-on in non-ESD time domain transient conditions may result in circuit or even printed circuit board components failure during functional testing. At the same time the event can be sensitive to the circuit board design due to significant change of the pulsed voltage waveforms caused by inductive and capacitive PCB load. In principle in case of ESD pulse regime a premature turn-on of the ESD device in snapback mode due to additional injection from the internal circuit components can be reversible and even favorable for the ESD circuit

operation itself. However, in actual operation with power supply voltage applied the turn-on of the ESD device may cause damage due to excessive power dissipation at high current.

A simplified circuit diagram with the switch pin (SW) is presented in Fig. 7.54a. The original circuit met all specified ESD requirements and was designed with the absolute maximum drain voltage of ~45 V for both switch (SW) and other separate control pins. The maximum operating voltage is specified at the level of ~40 V. The ESD clamp is based upon NDeMOS-SCR and designed for the triggering voltage of ~50 V targeting the lower ESD protection window limit of ~45 V. In the particular design identical ESD protection clamps have been used for protection of both the switch and the overvoltage protection pins. At the switch pin the topological separation of the clamp from the large power array was ~200 μm (Fig. 7.54b). The space is filled with the P-guard ring diffusion.

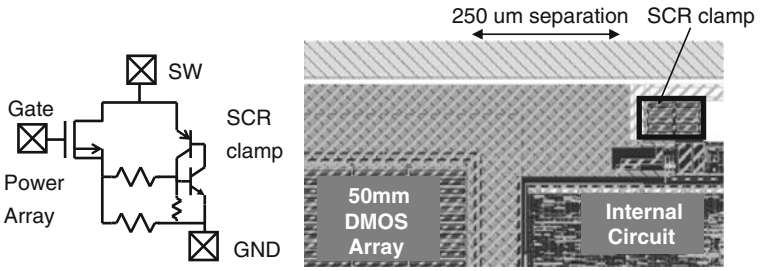


Fig. 7.54 Simplified application diagram for the part of the LED driver circuit (**a**) and layout of the power array and ESD clamp (**b**)

The observed application problem was related to accidental turn-on of the clamp in certain transient test regimes close to the absolute maximum limits.

The pulsed measurements demonstrate a significant change in the triggering voltage observed at the pin in comparison with the characteristics of the stand-alone clamp.

At switch pin to ground combination (SW–GND) TLP test results were different depending on the input pin voltage (VIN). With floating input pin the triggering voltage was measured as low as ~40 V (Fig. 7.55). Prior the triggering into snapback a significant difference of pin pulsed *I*–*V* characteristics was observed. Much higher ~3 A current level preceded the snapback mode (Fig. 7.55). This current is representative of the on-state conditions of the array connected in parallel to the pin.

Different VIN dc bias values resulted in different conditions at the power array gate in TLP transient regime. With VIN pin bias at high level the array gate potential is low and the clamp turn-on closely matches the stand-alone conditions. However, even in this case it is hard to conclude to what extent the internal avalanche–injection conductivity modulation mechanism of the clamp itself is involved. With

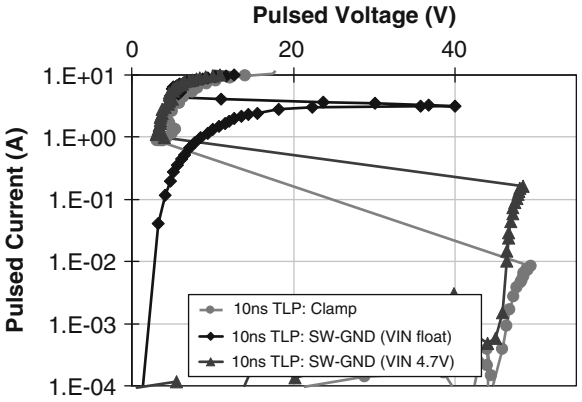


Fig. 7.55 Comparison of the TLP characteristics of the stand-alone ESD clamp with the characteristics at SW pin at different VIN voltage levels

bias applied to the input pin TLP characteristics of the SW–GND pin combination show the triggering voltage ~48 V much closer to the stand-alone clamp with low pre-trigger current. However, even in this case switching DMOS draws over ~150 mA current prior to the snapback mode (Fig. 7.55) indicating additional current path.

At floating VIN node the array gate to drain coupling of the switch transistor is high. This results in significant drain current and corresponding substrate current. In spite of the large guard-ring protected distance between the power array and ESD device the 10 V triggering voltage reduction is observed (Fig. 7.55).

Depending on the process, array design, and electrical regime the array substrate current can reach a level of ~30% of the drain current. The current path is realized between the array drain and any bulk p^+-diffusion in the vicinity of the arrays including P-guard ring of the ESD clamp. At the same time only a few milliamps of the current injected into SCR clamp is sufficient to trigger snapback mode.

Independent from triggering scenario the holding voltage at each pin is observed at the level similar to the stand-alone clamp. The 1–2 Ω higher resistance in this case can be attributed to extra ground bus circuit resistance. This fact confirms the same current path in the high-current snapback mode.

The control circuit pin to ground zap combination demonstrated TLP characteristics similar to the stand-alone device (Fig. 7.56).

This case demonstrated that despite layout isolation of SCR clamp, SCR clamp is still likely triggered by the parasitic current from internal circuit.

The substrate current in DeMOS array is acting similar to the base current of the parasitic NPN structure and provides lower turn-on voltage.

The following measures are usually considered to improve the design. For the low-impedance switch pins, in case of large switching LDMOS arrays $W > 30$ mm the circuit should rely on array self-protection. In this case corresponding design optimization of the power array may be required. An alternative approach may

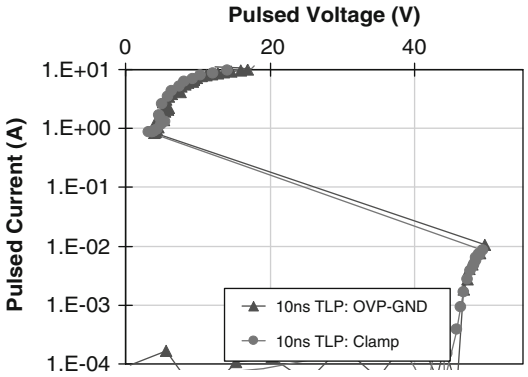


Fig. 7.56 Comparison of the TLP characteristics of the stand-alone ESD clamp with the characteristics at OVP pin

involve application of non-snapback high holding voltage ESD clamp based on lateral PNP or NPN device.

In case of small LDMOS arrays the gate clamp should be introduced to provide off-state for power array during ESD stress conditions and guarantee high breakdown voltage; however, this measure might be inadequate in fast transient test conditions.

For high-impedance control pins ESD clamp with pulsed triggering voltage significantly above absolute maximum can be used in combination with the two-stage protection approach. Typically this case is practical for the pins with large separation resistor.

Finally in application the design measures are related to appropriate PCB design that is minimizing the parasitic unductive components responsible for high-voltage overshoots in transient regimes.

7.6.2 Nepi–Nepi Transient Latch-Up Scenario

Another case of transient latch-up is typical for high-voltage processes. This latch-up current path can be realized either between two adjacent high-voltage pins protected by high-voltage ESD devices or the ESD device and the internal circuitry. The primary reason for latch-up path formation is parasitic NPN with variable base potential or simply current injection in ESD test conditions.

The example of parasitic NPN and corresponding current path is illustrated in Fig. 7.57 for the case of adjacent high-voltage control pin protected by NLDMOS-SCR ESD clamp and the power pin protected by high holding voltage lateral PNP ESD clamp. Both clamps can be referenced to the same power ground pin. However during ESD test the power ground node is floating. This in general creates variable conditions on the base of the parasitic NPN formed by Nepi regions of both structures and P-isolation ring.

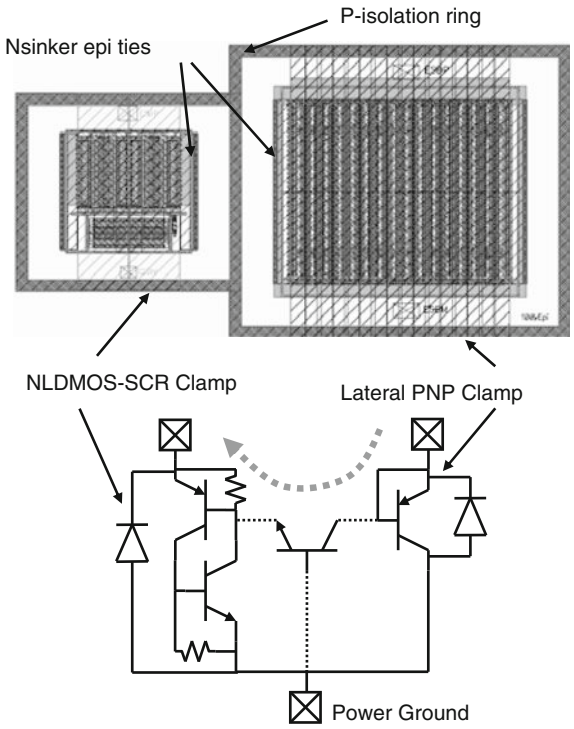


Fig. 7.57 Illustration of the alternative current path formed between two adjacent high-voltage pins protected by NLDMOS-SCR and lateral PNP clamps in BCD process technology

The originally expected current path should be realized in each direction through corresponding HV snapback clamp of positively biased pin and the reverse path diode of the negatively biased pin. However if two clamps placed with minimum isolation rules, an alternative current path is formed through the parasitic NPN structure, described above.

This alternative current path remains open for ESD current. Therefore, the final scenario depends upon the metallization routing and the internal circuit design. In a favorable case the new current path can simply positively contribute to the ESD protection level. However, in general an irreversible burnout may occur, since the real current path through the internal circuit is hard to predict Fig. 7.58.

Several important conclusions from this case study can be derived. First of all the high-voltage reverse path diode design in high-voltage clamps is critically important. For example, often for the pin protection itself the body diode of the clamp is sufficient. However, for the highlighted HV pin-to-HV pin combination the total voltage may exceed the parasitic NPN turn on voltage especially in fast transient mode, specifically due to excessive reverse path diode voltage drop.

A significant improvement of the body diode characteristics in high voltage lateral LPNP clamp in BCD process technology for reverse ESD protection can be accomplished by the following measures: enlarging n^+-base contact area, adding

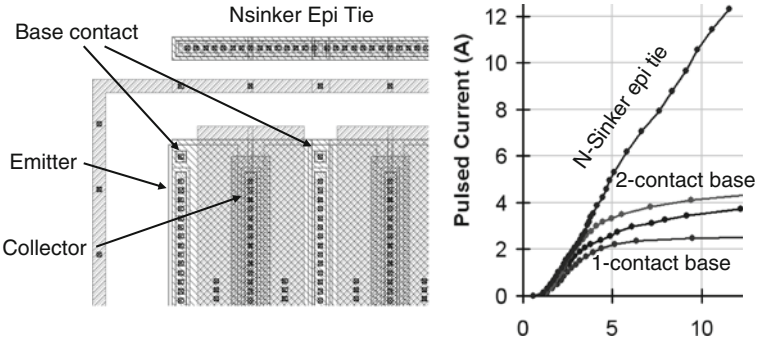


Fig. 7.58 Partial layout view of the Lateral PNP Clamp with the device regions and comparison of the negative TLP pulse characteristics for versions of the clamp with different N-base connection

base contacts or even by adding N-sinker epi tie to take an advantage of low buried N-layer resistance.

The effect of these measures is presented in Fig. 7.59 for the case of 100 V lateral PNP. All methods support 1.5 A current (>2 kV HBM) at less than 3.5 V voltage drop. The highest current and lowest resistance achieved with N-sinker as expected, which also improves the forward TLP *I–V* characteristics.

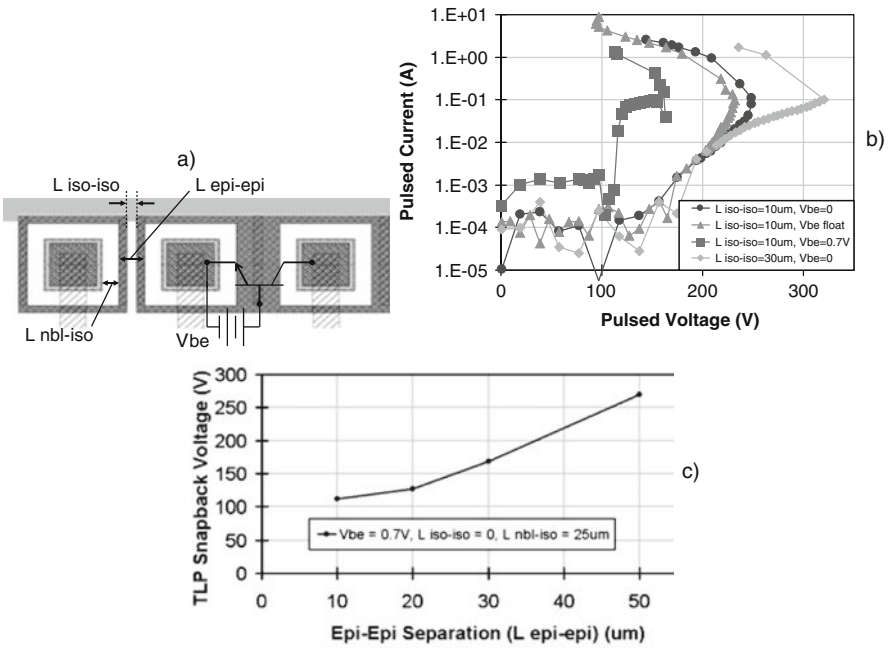


Fig. 7.59 Layout view of experimental structures for epi-to-epi isolation rules evaluation (**a**), measured TLP characteristics for different spacing and isolation ring bias conditions (**b**), and experimental dependence of the critical snapback voltage upon epi region separation (**c**)

Meanwhile, the main approach to suppression of the parasitic current involves introduction of ESD layout guidelines for spacing limits between two epi- regions connected to the high voltage. A simple three-terminal experimental device structure to study the current path through parasitic NPN can be created and characterized to determine the rules for different high-voltage epi regions isolation in ESD conditions. According to the experimental data the critical voltage is a function of the isolation region bias and the spacing between the epi-regions that represents the base width.

7.6.3 CDM Case of the High-Voltage Pin Protection

Typically CDM ESD pulse is not of a major concern for small pin count analog circuits due to smaller parasitic capacitance of the package. However, in some cases CDM pulse directly impacts even small pin count high-voltage pins and special measures should be taken.

Specific example below demonstrates the case of CDM failures in avalanche diode-referenced CDM clamp (Fig. 7.60). The clamp is used to protect high voltage-tolerant under-voltage lockout pin (UVLO). An external resistor divider from the system input voltage sets the under-voltage turn-on threshold. An internal very low current source provides hysteresis. This pin can be used for remote shutdown control. The ESD-related specific characteristic of this pin is a very low internal circuit load.

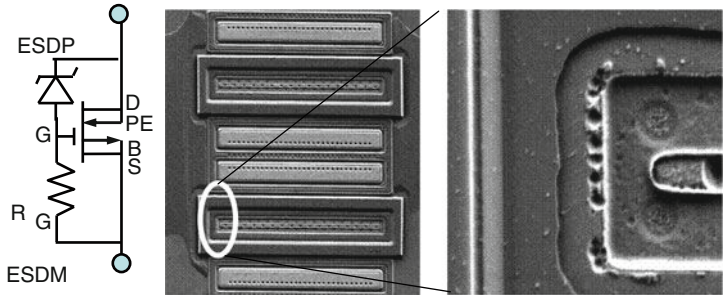


Fig. 7.60 Simplified circuit diagram for NLDMOS-SCR clamp with high-side reference and failure analysis photos of the damaged gate–source region of the NDeMOS-SCR clamp device

In a power product several pins have been protected by this clamp; however, only the pins with low internal load, similar to UVLO, failed during ESD tests. Failure analysis revealed gate–source region oxide damage.

Consequently the physical mechanism of the CDM failure is concluded from failure analysis results to be gate oxide breakdown under CDM stress. The dominant current path is through the high-side avalanche diode capacitance. To eliminate

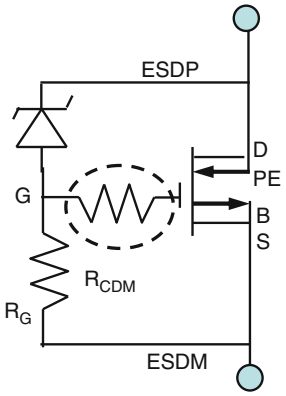


Fig. 7.61 An additional gate resistor R_{CDM} design measure to eliminate CDM failure of the high-side-referenced NLDMOS-SRC ESD snapback device

this undesired effect an additional resistor to protect the gate of the ESD protection clamp is proposed (Fig. 7.61).

The additional experimental analysis for this case is presented below to explain the process in the high-side-referenced NLDMOS SCR clamp under CDM stress. The difference of the clamp characteristics in the CDM time domain can be revealed by comparing the 100 ns standard TLP and 2 ns very fast TLP (vfTLP) measurements. The difference is observed for both the drain–source and source–drain pulsed *I–V* characteristics.

In case of reverse clamp path the clamp diode speed is not sufficient to clamp the reverse path voltage low. For the particular package the 750 V CDM pulse roughly corresponds to 5 A peak current level.

Comparison of the TLP and vfTLP (Fig. 7.62) shows significant difference in the clamping voltage due to the difference in the measurement time domain.

Additional comparison using vfTLP measurements is made for the clamps with and without high-side reference component. The result demonstrates much lower maximum current for the clamp with high-side reference component (Fig. 7.63).

Further insight on this CDM failure is obtained using mixed-mode simulation analysis. Based on simulation results it can be demonstrated that in the case of high-side reference not only reverse, but even forward current path creates gate voltage overstress.

Figure 7.64 provides comparison of the positive CDM pulse simulation for four representative cases of the circuit: (i) with disconnected high-side avalanche diode reference; (ii) with connected diode reference; (iii) the effect of the additional gate protection resistor; and (iv) the effect of the additional capacitive load parallel to the pin represented by reverse path diode. For positive CDM current direction the presence of the reference components results in three times higher gate peak voltage that explains the gate oxide damage (Fig. 7.64c). This undesired effect can be

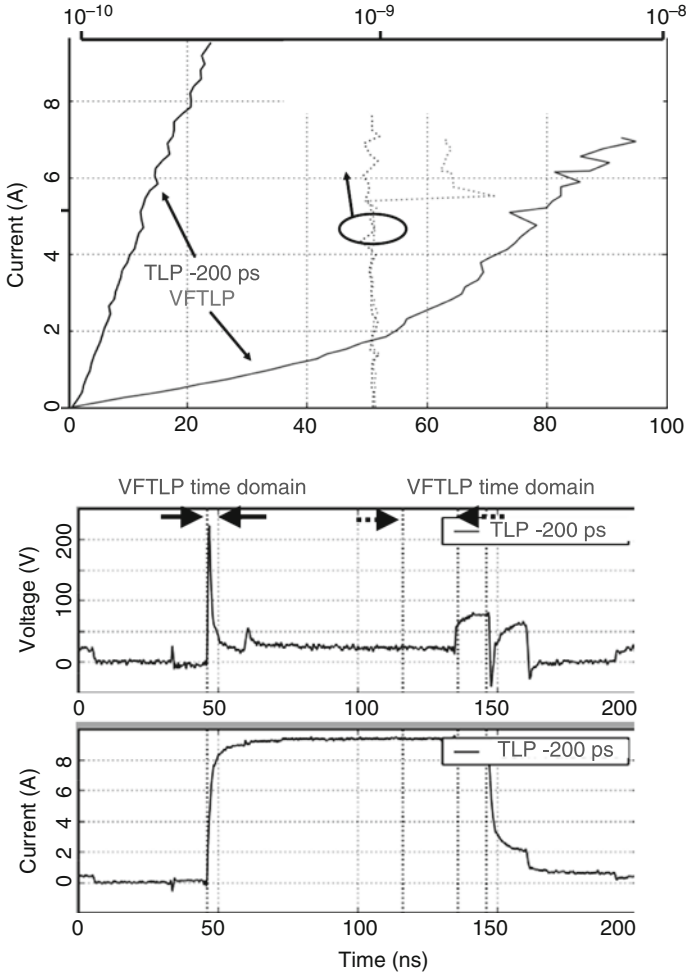


Fig. 7.62 Comparison of the 2 ns vfTLP and TLP characteristics and illustration of the different time domain measurement windows for the case of TLP and vfTLP

significantly reduced both by the additional gate protection resistor and by addition of the load (Fig. 7.64c).

A different scenario is observed for negative stress. In this case the effect on the reverse path diode or the body diode provided by internal circuit components is critical. In case of 100 ns TLP or HBM time domain the parasitic high-voltage diode formed by the drain and bulk Pbody or P^+-guard ring connection is usually sufficient for reliable current path. A different behavior is observed in case of very fast CDM event transient pulse. The waveforms for 500 V positive CDM stress with and without resistor and with additional reverse path diode load are compared in Fig. 7.65.

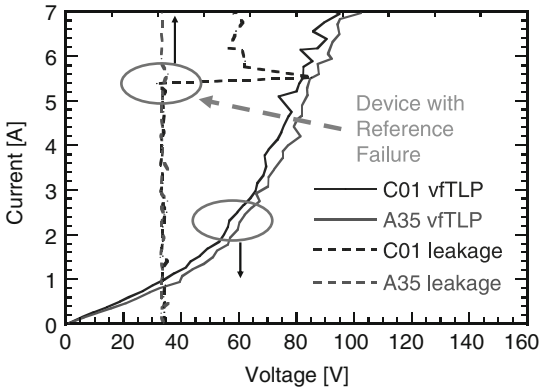


Fig. 7.63 Comparison of the 2 ns vfTLP characteristics for the clamp with and without reference device

A significant overstress of the gate in CDM time domain is observed if the high-side reference component is connected and there is no load. An improvement is achieved by adding the gate resistor as well as in case of extra load provided by body diode.

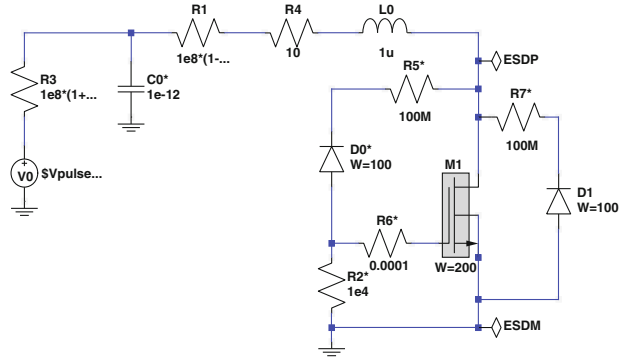
7.7 Summary

In this chapter, applications of ESD network principles and clamps are discussed within the scope of different power circuits.

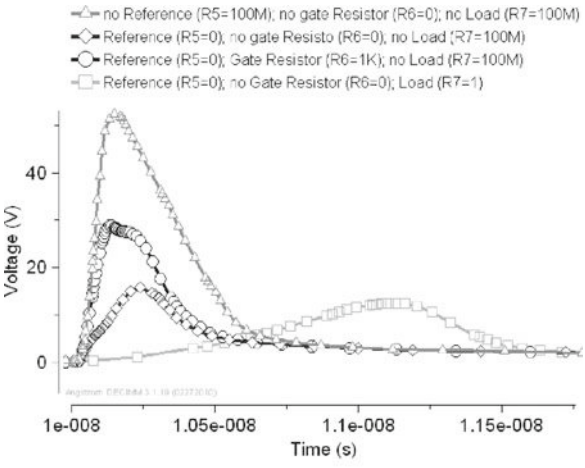
The specifics of power circuits significantly impact the ESD protection strategy. Fast-switching and high-voltage pins, together with the lack of an ESD protection window due to power-optimized output devices, create significant challenges and often require circuit-dependent, custom ESD approaches. Self-protection aspects of the integrated power devices in complex, multiple power domain networks are hard to simulate and require substantial product-level experimentation.

In spite of similar basic principles for ESD protection network design and ESD clamp solutions, a new characteristic feature of the power circuit consists in the multiple current paths that can be realized during ESD event.

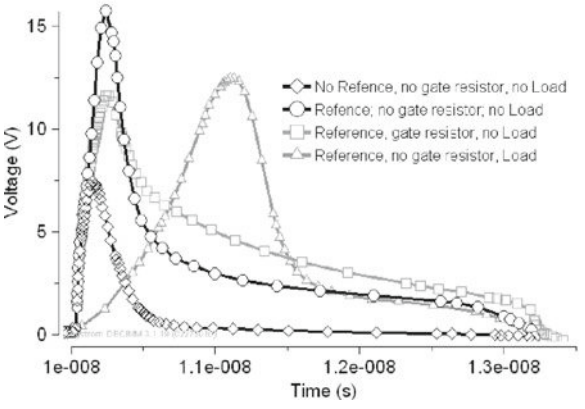
In the case of fast transient pins and output devices with pulsed absolute maximum voltages too close to operational limits, a significant effort is put into focused design and selection of clamps with a precise voltage reference and a low dV/dt triggering effect. The possibility of transient latch-up on transient pins or in case of a hot plug-in circuits requires countermeasures. Among them are the design and application of rather challenging high holding voltage solutions (especially in the case of high-voltage circuits) and utilization of the self-protection capabilities of the internal circuits. At the same time, realization of the two-stage ESD protection network becomes rather challenging due to the impossibility of putting a separation



a)



b)



c)

Fig. 7.64 Mixed-model simulation circuit with positive 2000 V CDM pulse (**a**) and pulsed waveforms observed at the drain (**b**) and at the gate (**c**) of the ESD device for the cases with and without high-side reference, gate protection resistor, and reverse path diode load

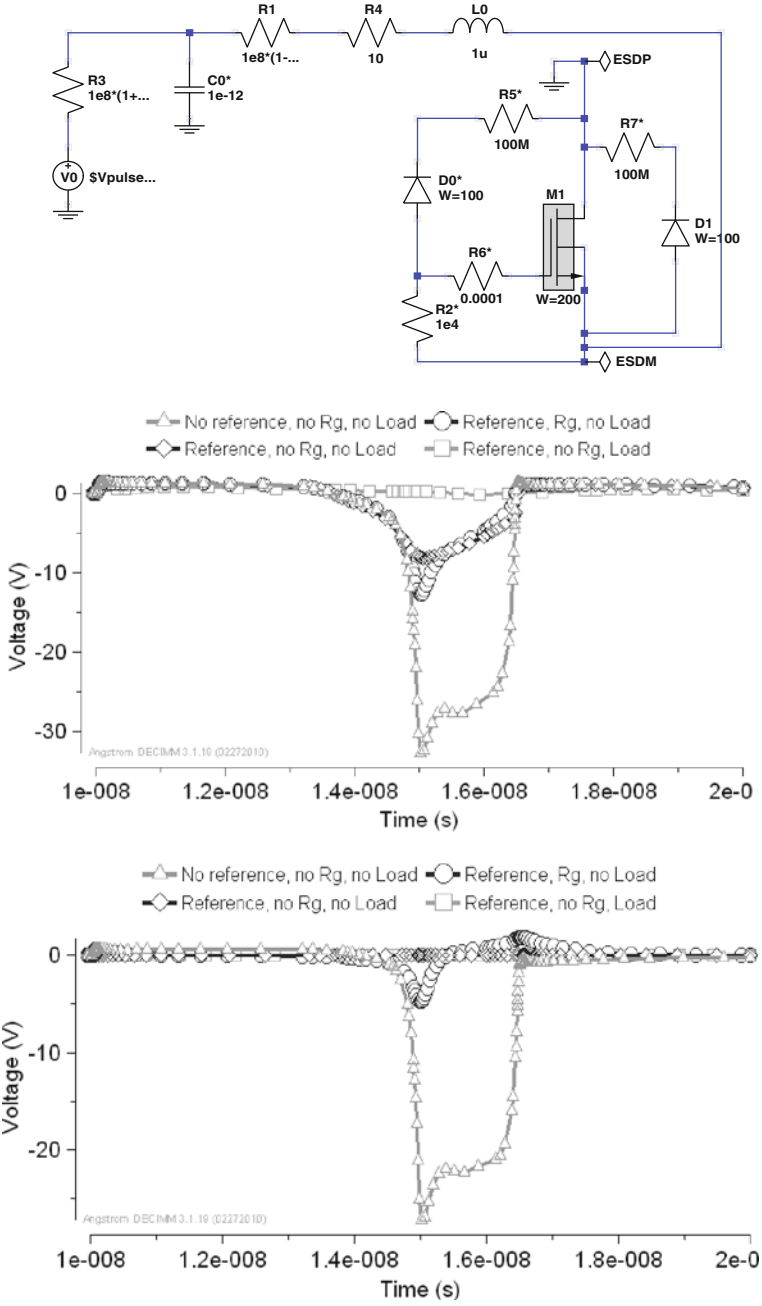


Fig. 7.65 Mixed-mode simulation circuit with negative CDM pulse (**a**) and pulsed waveforms observed at the source (**b**) and at the gate (**c**) node of the ESD device (relative to the ground) for the cases with and without high-side reference, gate protection resistor, and reverse path diode load

resistor on the power node. As a result, the strategy for protection changes from rail based protection to individual device protection.

A general methodology of implementing appropriate ESD protection contains several steps. It depends upon whether the ESD protection network is created for an existing, active design, or, conversely, a co-design of the internal voltage regulator circuit, the ESD protection, and even the components can be accomplished. The latter is apparently preferable.

The fist step of the methodology is to identify the pin types: control, power, switch, boost, the functionalities of feedback pins, voltage tolerance. Transient signal, and other specification parameters determine the most suitable network design and the most appropriate clamps.

The second step requires a significant amount of "navigation" from each pin inside the circuit block hierarchy in order to reveal the most exposed devices that will provide an alternative current path. The following step is focused on estimation the reversible current level in ESD conditions which can be supported by the internal circuit without any damage to the active devices or interconnects. To achieve a suitable level of confidence in the estimation, understanding of both the pulsed SOA of the components and the circuit operation is critical.

These two steps usually require substantial effort and experience with similar products. They involve a number of assumptions and provide relatively low initial accuracy of estimation for unique circuit design cases. However, this is the most widely used approach among industry experts. Part of the needed experience is illustrated with the case studies presented in this chapter.

However, the most effective way is to combine this valuable experience-based empirical approach with simulation tools. Examples of the mixed-mode and compact model simulations with ESD snapback models have been presented across this chapter and described extensively in Chapter 5.

The following steps involve choosing the ESD protection network components and selection of clamps with the most appropriate characteristics, as well as increasing the self-protection capabilities.

When the ESD protection network schematic is finalized, finding an appropriate ESD layout is the next critical step. In addition to the obvious backend layout design that supports the ESD current level at a low voltage drop, more thorough work is required to eliminate possible cross-talk (due to the injected current into the substrate) between the internal circuit and the ESD clamp. This is usually achieved by both the guard ring structure and an appropriate arrangement of the ESD clamp on the chip. For example, in most cases, the high-voltage ESD clamp should be sufficiently separated from the power array to avoid the false triggering by the high substrate current provided by the array. Other layout-related efforts are focused on the elimination of the parasitic coupling between the metal layers that can provide undesirable gate coupling of the components.

For integrated power arrays, the self-protection capability is a layout-dependent parameter [136]. Recent research work [113] demonstrates the possibility of estimating the self-protection operation of power arrays using a 2.5D approach to extract an equivalent circuit of the array combined with simulation [137] and visualization [138] tools.

When the circuit and layout design work is completed, the incoming test results are the most important factor for further understanding of the ESD network and circuit coupling. If reliability improvement steps have to be taken because of product failures under ESD test, the partition test, combined with reprocessing-level FA results, becomes very important. Usually, the failure location is repeatable and the major challenge is in creating an appropriate solution. Product-level validation of the alternative solutions is usually a rather expensive approach. This is why both empirical analysis and numerical evaluation of the proposed solution may greatly contribute to the level of confidence in its success.

DECIMM^{TM} Simulation Examples for Chapter 7

To download a trial version of the numerical simulation software and request an electronic license key please visit http://www.analogesd.com
To download libraries with simulation examples for this chapter please visit http://www.analogesd.com/Chapter7.html
List of examples is subject to change.

Example 7.1 Output Stage of Buck DC–DC Voltage Regulator

Library Name: Examples7_Power_Management_Analog_Circuits
Project Name: E7.1a_5 V_Synchronous_Buck_No_Bond_Wire_Inductance
Project Name: E7.1b_5 V_Synchronous_Buck_Normal_Operation
Project Name: E7.1c_5 V_Synchronous_Buck_Transient_Latchup_Case

This set of examples presents case studies for the simplified version of the output stage of the synchronous 5 V buck DC–DC converter. In one example, the bond wire inductance effect is shown to produce a high-voltage transient at the power pins. In normal operation regime, depending on circuit parameters, this may result in a transient latch-up event due to triggering of the NMOS ESD clamp into snapback. The mixed-mode simulation examples for the output stage of synchronous DC–DC converter can also be used to understand the principles of operation of the voltage regulator and the effect of the physical parameters of the finite element components on the output stage characteristics, including switching speed and losses (Fig. E7.1).

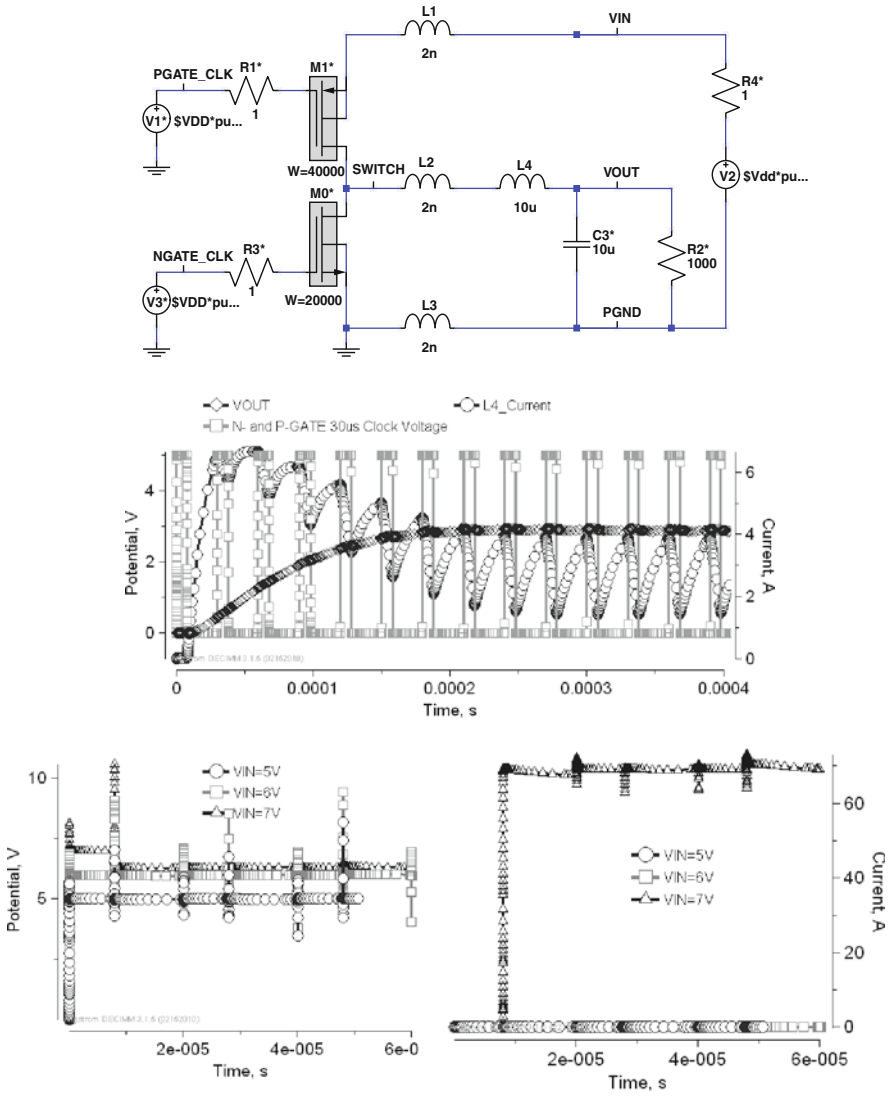


Fig. E7.1 Mixed-mode simulation circuit for simplified output stage of 5 V DC–DC converter including bond wire inductances and simulation results for normal operation regime; mixed-mode circuit for transient latch-up simulation and results of analysis demonstrating induced parasitic snapback in the ESD protection clamp device

Example 7.2 5 V Boost DC–DC Converter and Transient Latch-Up

Library Name: Examples7_Power_Management_Analog_Circuits
Project Name: E7.2a_5 V_Boost_HNMOS_Normal_Operation
Project Name: E7.2b_5 V_Boost_HNMOS_Transient_Latchup

This set of examples presents case studies of the simplified version of the output stage of the asynchronous 5 V boost DC–DC converter with the switch pin protected by a snapback NMOS ESD clamp. The example demonstrates both the principle of voltage regulation and possible transient latch-up event due to snapback of the NMOS ESD clamp (Fig. E7.2).

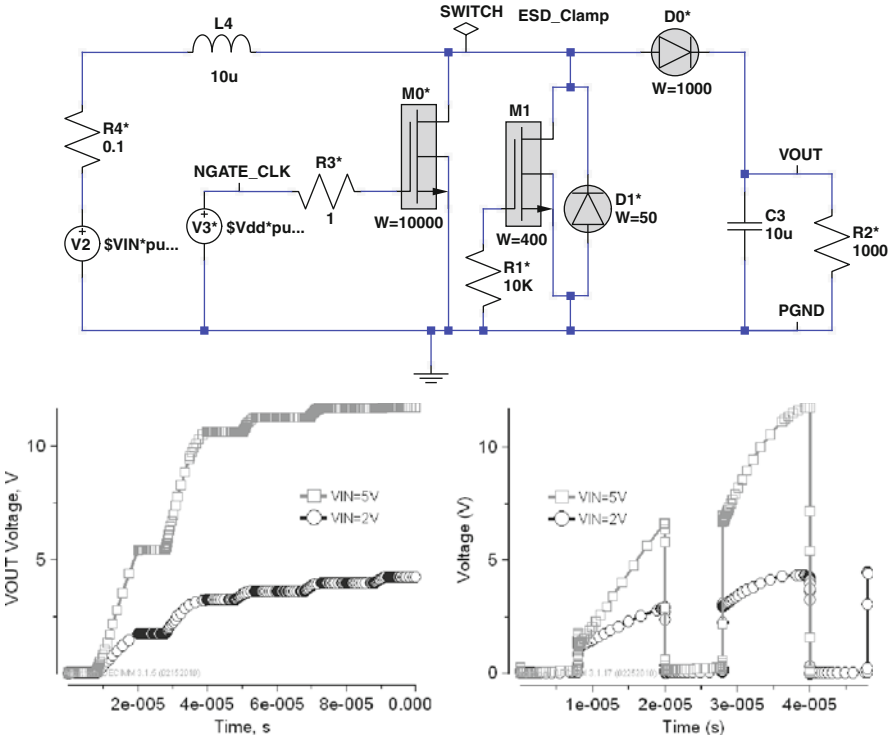


Fig. E7.2 Mixed-mode simulation circuit for the simplified output stage of the 5 V DC–DC boost converter and simulation results for the normal operation regime showing output and switch voltages for different VIN voltage levels

Example 7.3 High-Voltage Boost Output Stage

Library Name: Examples7_Power_Management_Analog_Circuits
Project Name: E7.3_20 V_Boost_NLDMOS

This example provides a starting point for mixed-mode analysis of a simplified version of the output stage of the asynchronous 20 V Boost voltage converter (Fig. E7.3). An optional high-voltage ESD protection clamp can be added to analyze parasitic effects during ESD stress and under normal operation.

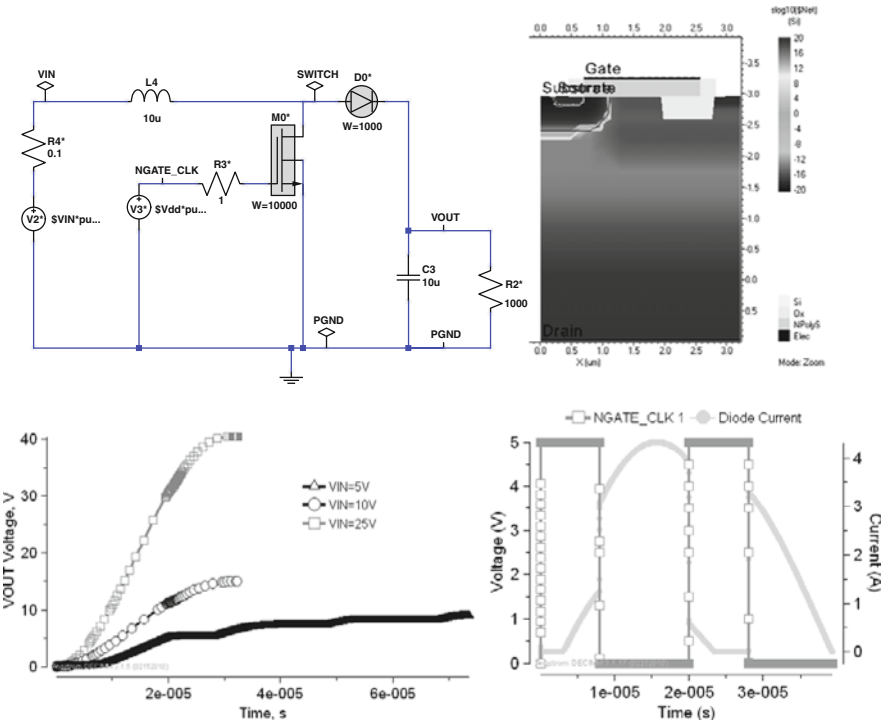


Fig. E7.3 Mixed-mode simulation circuit for a simplified output stage of the 20 V DC–DC boost converter, simulation results for the normal operation regime showing output voltage VOUT at different VIN voltage levels, gate drive voltage and diode current waveforms

Example 7.4 100–600 V Boost Output Stage with Vertical DMOS and IGBT

Library Name: Examples7_Power_Management_Analog_Circuits
Project Name: E7.4a_100 V_Boost_VDMOS
Project Name: E7.4b_100 V_Boost_IGBT

These examples present mixed-mode analysis of a simplified version of the output stage of the 100–600 V boost voltage regulator. The output stage is assembled using parameterized vertical DMOS and IGBT devices that represent discrete components (Fig. E7.4).

Example 7.5 Power Array with Gate Clamp Example

Library Name: Examples7_Power_Management_Analog_Circuits
Project Name: E7.5_Gate Clamp

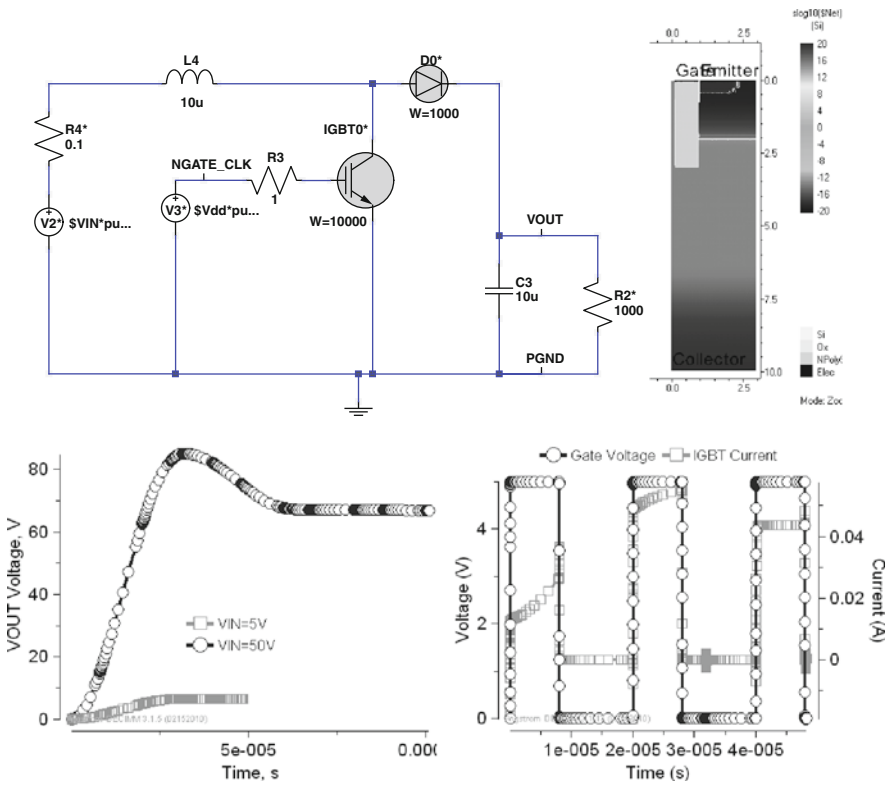


Fig. E7.4 Mixed-mode simulation circuit for a simplified output stage of the boost converter and simulation results for the normal operation regime showing output voltage VOUT at different VIN voltage levels, gate drive voltage and VDMOS drain current waveforms

The example presents a version of the so-called gate clamp implementation. The purpose of the gate clamp is to ensure a relatively low gate bias of the NMOS array in the ESD pulse time domain, but enable switching of the device in normal operation mode. A simulation with two HBM circuits is performed in parallel to demonstrate the transient gate voltage of the power device with and without the gate clamp (Fig. E7.5) and the higher drain voltage in the circuit with the gate clamp. Additional optional snapback clamps can be added in parallel to the power array to observe the alternative ESD current paths through the clamp and though the array.

Example 7.6 Serial Data Line Pin Case

Library Name: Examples7_Power_Management_Analog_Circuits
Project Name: E7.6_SDA_pin

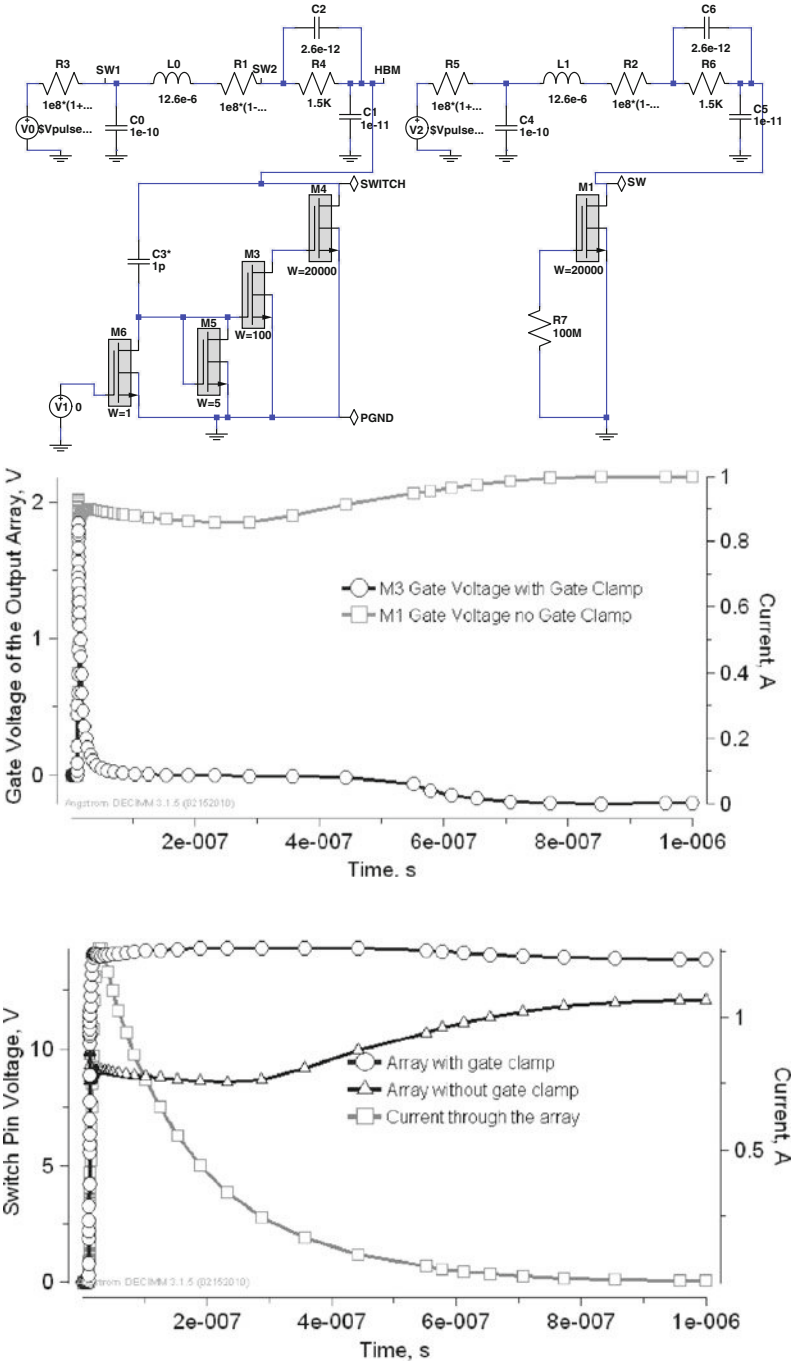


Fig. E7.5 Mixed-mode simulation circuits for power array with and without gate clamp and simulated gate and drain voltage waveforms

This example represents the case of the serial data line pin ESD protection with a snapback NMOS clamp. Different current paths through the clamp and the internal NMOS device can be observed in the simulation, depending on the passive component values and active device parameters (Fig. E7.6).

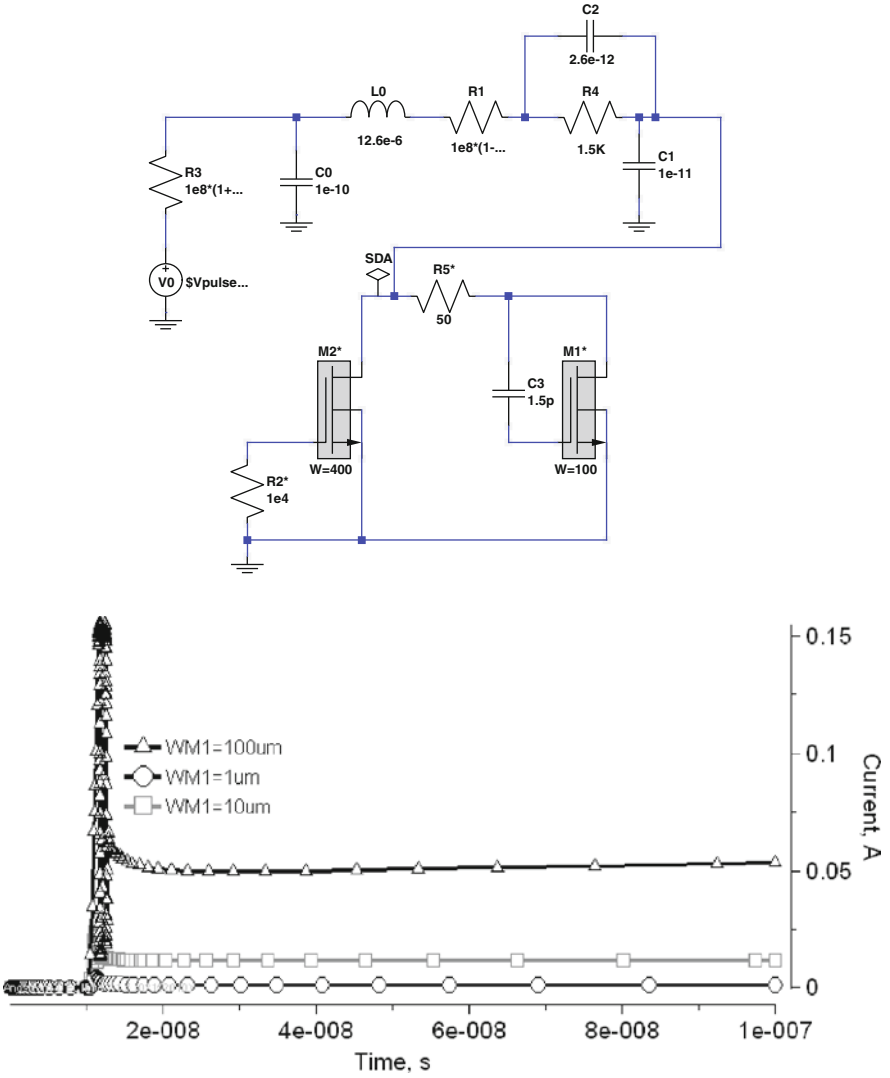


Fig. E7.6 Mixed-mode simulation circuit and transient current through the internal circuit NMOS M2 for different width of the protection device M1

Chapter 8
System-Level and Discrete Components ESD

This final chapter presents material on a rather cross-disciplinary subject related to system-level ESD robustness. System-level ESD requirements are defined by different standards and specifications than component-level ESD requirements. Component-level ESD standards are specified to ensure ESD robustness of the integrated circuits and components during manufacturing and handling inside the controlled, protected ESD environment until the components are incorporated into the system. ESD specifications on the component level are circuit specific and package specific, practically until the component is mounted on the print circuit board. It is expected that when the component is mounted in the properly designed system, the system's design will guarantee the absence of events that result in stress of the components above the absolute maximum limits, even when the system itself experiences an ESD event.

System-level ESD robustness is required to ensure robustness of the whole system during system manufacturing and operation. This can be achieved both by different system design architecture that involves transient voltage suppressors and by implementation of the on-chip ESD protection for the pins. The focus of this chapter is specifically on the on-chip system-level protection.

The chapter is composed as follows. First, Section 8.1 provides an introductory portion of the material related to the understanding of the standards and ESD system specifics. Section 8.2 brings the human–machine model approach that enables on-chip system design for ESD clamps and protection networks. In Section 8.3, several application cases for on-chip ESD protection are discussed to demonstrate on-chip protection peculiarities and the general approach. A solution to one of the particular cases of system requirements related to hot swap and plug-in is discussed in Section 8.4. Finally, in Section 8.5, the chapter is complemented by a discussion of the ESD robustness of discrete components.

V.A. Vashchenko, A. Shibkov, *ESD Design for Analog Circuits*,
DOI 10.1007/978-1-4419-6565-3_8, © Springer Science+Business Media, LLC 2010

8.1 System-Level Specifications and Standards

8.1.1 Meaning of ESD Robust System

ANSI/ESD S20.20 is an ESD control program that provides a systematic approach to protecting components and equipment susceptible to ESD damage starting from HBM discharges of >100 V. A set of standards related to the measurement of static control in footwear, packaging, flooring, conveyor belts, ionizers, etc. However, it is difficult to maintain this level of control 100% of time and everywhere due to the cost associated with these measures.

The component-level ESD standards protect the ICs from damage within a static controlled environment. These standards assume basic ESD control practices that are typical for a 2 kV human body model, a 200 V machine model, and a 750 V charge device model. This applies to IC's before they are mounted on a board.

Thus, the focus of ESD protection shifts from chip hardening to system hardening. The previously mentioned CDM, MM, and HBM standards apply to chip protection, but do not guarantee system level protection. The end system is tested according to system-level standards with an "ESD gun." For example, the surface of the device is divided into a grid. Each grid space is zapped a minimum of 10 times (Fig. 8.1a). Failure can be catastrophic damage, latch-up (device requires power-off to reset), or "soft" damage, for example, a display flicker.

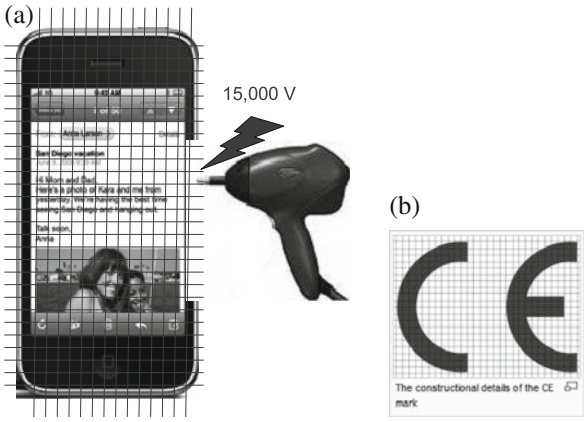


Fig. 8.1 System testing wit ESD gun (**a**) and CE mark image (**b**)

In principle, successful system design requires understanding beyond the chip level. In particular, ESD is not just a handling issue. For system-level protection design, it is important to know what circuitry is connected on the board to the chip in the final system design, if the particular current path is directly connected to the external system pins, and what kind of connection and cable is in place. More details may be required to identify if the connection can be probed or touched during system repairs or maintenance. Thus, understanding of the relevant system (ESD) qualification standards can directly affect the design.

ESD robustness is one of the quality features of the system. At the same time, the system is the final application of an integrated circuit component. A robust ESD system ensures reliable operation during and after ESD stress.

8.1.1.1 CE Mark

Depending on the criticality of the application, different levels of functionality have to be maintained during or after ESD stress. Passing appropriate standard ESD tests is a legal requirement for systems that udergo corresponding certification procedures. For example, the IEC 61000-4-2 standard has to be met to obtain the CE certification mark. Since 1996, all medical consumer electronics and many other systems must conform to the 89/336/EC (EMI and ESD) and display the CE mark to be sold in Europe.

The CE mark (Fig. 8.1b) is a label affixed to a product to show its compliance with the electromagnetic compatibility (EMC) directive and with the appropriate safety directive: the machinery directive for the safety of machines or the low-voltage directive for the safety of electronic products. The manufacturer is responsible for selecting the safety directive most appropriate for his product. Many products include both the electrical and mechanical components.

The standards hierarchy for EU includes Product or Product Family: tests, test level, failure criteria; Generic EMC: simulators, test setups, procedure, if a product standard does not exist; Basic EMC: simulators, test set-ups, procedure.

8.1.1.2 Basic EMC Standards

Practically every manufacturer in the system development business must perform tests to satisfy the following Basic EMC Standards:

ESD immunity	EN 61000-4-2
EFT – electrical fast transients	EN 61000-4-4
Surge immunity	EN 61000-4-3/6
RF emissions	Various
Harmonic/flicker tests	EN 61000-3-2/3

Recently, European Norm (EN) naming conventions have been used for existing standards: EN61000-4-x = IEC 61000-4-x, etc.

The object of the IEC 61000-4-2 standard is to establish a common reproducible basis for evaluating the performance of electrical and electronic equipment when subjected to electrostatic discharges; typical waveform of the discharge current; the range of test level; test equipment; set-up; test procedure.

8.1.1.3 Automotive Industry Standards

Automotive industry applies the ISO 10605 standard. Functional Status according the ISO 10605 includes the following classes:

– Class A: all functions of a device or system perform as designed during and after exposure to interference.

– Class B: all functions of a device/system perform as designed during exposure; however, one or more may go beyond the specified tolerance. All functions return automatically within normal limits after exposure is removed. Memory functions shall remain class A.
– Class C: one or more functions of the device or system do not perform as designed during exposure, but return automatically to normal operation after exposure is removed.
– Class D: one or more functions of the device or system do not perform as designed during exposure and do not return to normal operation until exposure is removed; the device or system is reset by a simple operator/use action.
– Class E: one or more functions of the device or system do not perform as designed during and after exposure and cannot return to proper operation without repair or replacement of the device or system.

8.1.1.4 IC and System-Level Comparison

It is important that the tests and protection methods are applied specifically to ESD robust systems (like IEC 61000-4-2), rather than components. In principle, the components formally do not necessarily need to meet any special requirements beyond the safe packaging and handling if the system design is compliant with the system-level specifications. Thus, for components, there are other ESD discharge models that might be more relevant: HBM, MM, charged device model (CDM), charged board model (CBM), and the cable discharge event (CDE) [139].

However, the trend set by the system designers is to elevate the passing level for component ESD tests on the chip in order to help system-level ESD design.

Manufacturing and application cycles of the IC component can be subdivided into the IC component-level ESD and system-level ESD (Fig. 8.2). During the component-level ESD, the IC is always handled inside the ESD protected area (EPA), where ESD protection methods and design are implemented in a well-controlled environment. In this environment, presence of ESD events with pulse amplitude above the component-level ESD (2 kV HBM, 200 V MM, 1 kV CDM) is practically excluded. In reality, the levels of ESD are much lower. This is why reduction of the component-level minimum ESD requirements to 1 kV HBM and 100 V MM is recently in discussion.

Thus, practically, on-chip ESD protection compliant with several hundred volts HBM guarantees safe handling. Protection realized between any pin combinations covers the life cycle from backend manufacturing to board mounting inside EPA.

Outside EPA, the protection methods and design of the system include several measures. For example: shielding, pre-running ground, application of the system-level ESD components, for example, transient voltage suppressors (TVS), to achieve the desired system-level specification.

In general, system-level ESD protection is necessary for satisfying the corresponding much higher levels of ESD currents, in comparison with the package-level requirements. At the same time, properly designed ESD protection on the packaging level can protect components even outside EPA. Usually, the enhanced protection

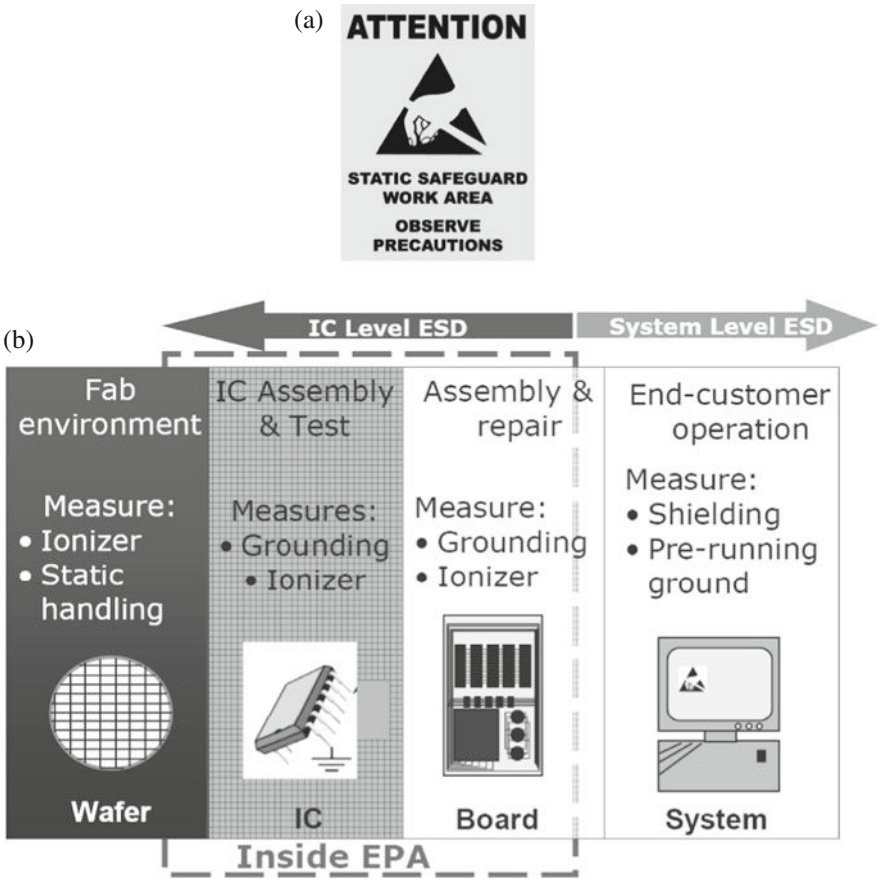


Fig. 8.2 Typical sign to designate ESD-protected area (EPA) (**a**) IC and system-level ESD protection and measures (**b**) [139]

design applies only to the very few exposed pins with system-level requirements, rather than to the whole component. In this case, IC-level ESD protection design for the non-exposed pins does not influence system-level ESD robustness.

A typical system-level ESD exposure during assembly can include different scenarios. For example, in a car assembly, a discharge may occur when a cable is plugged into a control unit. In a mobile assembly, a discharge is typical when boards are handled or sub-boards are connected to the main board. Typical system-level ESD exposure situations during end-user operation usually involve antenna pins overstress; picking up energy of an ESD discharge in a nearby surrounding area; discharge of cables and connectors during plug-in or wiring; damage of the battery supply pin during changing of batteries; discharge inside a system due to board or housing design flaws.

In effect, any metallic part which can be touched by the end-user and which is close to the IC pins or the board can expose system to an ESD event. System-level ESD, according to IEC 61000-4-2, mimic a discharge to a system (outside an EPA) from a handheld metallic tool.

For the system level ESD, direct discharge to "open" connectors is excluded. For communication systems, discharges through cables and equipment may be real threats (CDE – cable discharge event). Examples include USB, FireWire interface in PC/camera/peripherals; Ethernet (cables may be very long), BNC cables; automotive wiring harness. Different discharge types may contain an "initial discharge" that corresponds to the connection of the charged cable to the pins, a "secondary discharge" that corresponds to a change in the connection, for example, a short circuit at the other end of the cable, or a combination of discharge from both cables and devices [140].

8.1.2 System-Level ESD Pulse and Model

8.1.2.1 System-Level ESD Test for ICs

The application board capacitor has already been emphasized for the case of the FDP-link (Section 6.3.3). Thus, at system-level IC testing, it is important in principle to have a special application board specifically for this type of testing. The role of the application board is to mimic the final system board size and the ground planes. In addition, special modifications to the board are often required to allow zap access for the pins, due to standard system-level gun dimensions. In most cases, only some specific pins are required to undergo system-level tests while other pins are not subject to such requirements. One of the challenges is correctly interpreting these requirements: what ports are to be tested; what is the system board design.

The board design itself can also help pass the system-level test. For example, in addition to the decoupling capacitors effect, the board inductances will attenuate the peak current (Fig. 8.3).

Typically for on-chip system-level protection there is an area and performance trade-off. On chip protection provides an advantage in reduction of component count, but at the same time creates an overhead in silicon area, and complication in test/qualification. The complications may require running several learning cycles to correctly target the "ESD window." Thus, component and system tests must be carefully specified in advance. The real target is the result in the end system.

8.1.2.2 IEC ESD Pulse Waveforms

System-level ESD pulse has two-peak waveform (Fig. 8.4a, b) according to IEC 61000-4-2. On the physical level, the first peak is caused by a metallic part that touches the system-level pin. This peak has a fast rise time and high amplitude, but is short due to the low resistance of the metallic part and limited charge. The first peak is followed, with corresponding delay, by a discharge that physically represents

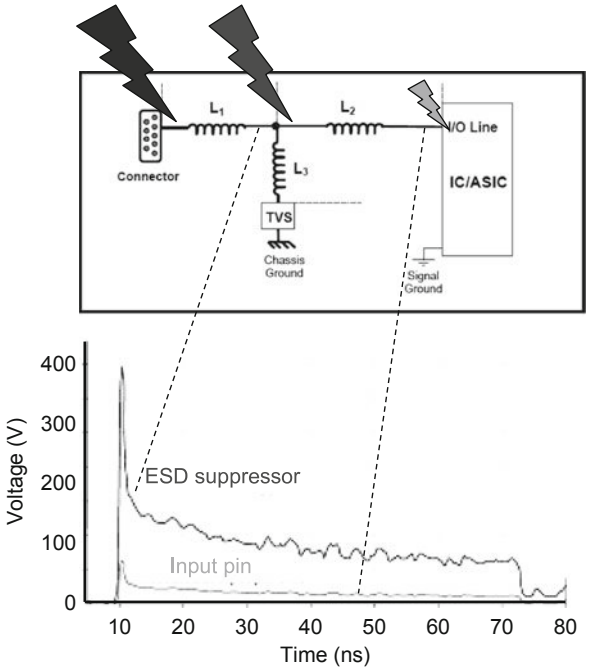


Fig. 8.3 Various inductance components on representative board trace and comparison of the measured waveforms at ESD suppressor and at input pin of an IC

a charged body of an end-user or a tool. This peak has lower amplitude due to higher resistance versus the metallic part and a corresponding delay and width, because the object discharged is much larger.

Standard documents usually define a waveform, not the equivalent circuit. For the IEC 61000-4-2 pulse, the waveform is defined by several representative points of current amplitude during the transient event. For the first peak, the specific current per 1000 V of the precharge level is $I_{1\text{st peak}} = 3.75$ A/kV; the second specific current level per precharge level is defined after 30 ns from the first peak: $I_{30\text{ ns}} = 2$ A/kV; the next current level is $I_{60\text{ ns}} = 1$ A/kV. In addition, the rise time for the first peak of the pulse is defined within t_{rise}=100-300 ps and full width ≈ 1 ns.

To replicate the IEC 61000-4-2/ISO 10605 standard, one of the main components of the equivalent tester circuit can be roughly represented by the $R_{\text{sys}}C_{\text{sys}}$ network with parameters 330 Ω and 150 pF. However, such a network will not provide a desired waveform by itself. A more accurate equivalent circuit is discussed in Section 8.2 for the human machine model (HMM).

A comparison of HBM and HMM pulses of the same amplitude is presented in Fig. 8.4b. As can be seen, the differences between the waveforms are in both the pulse shape and peak current level.

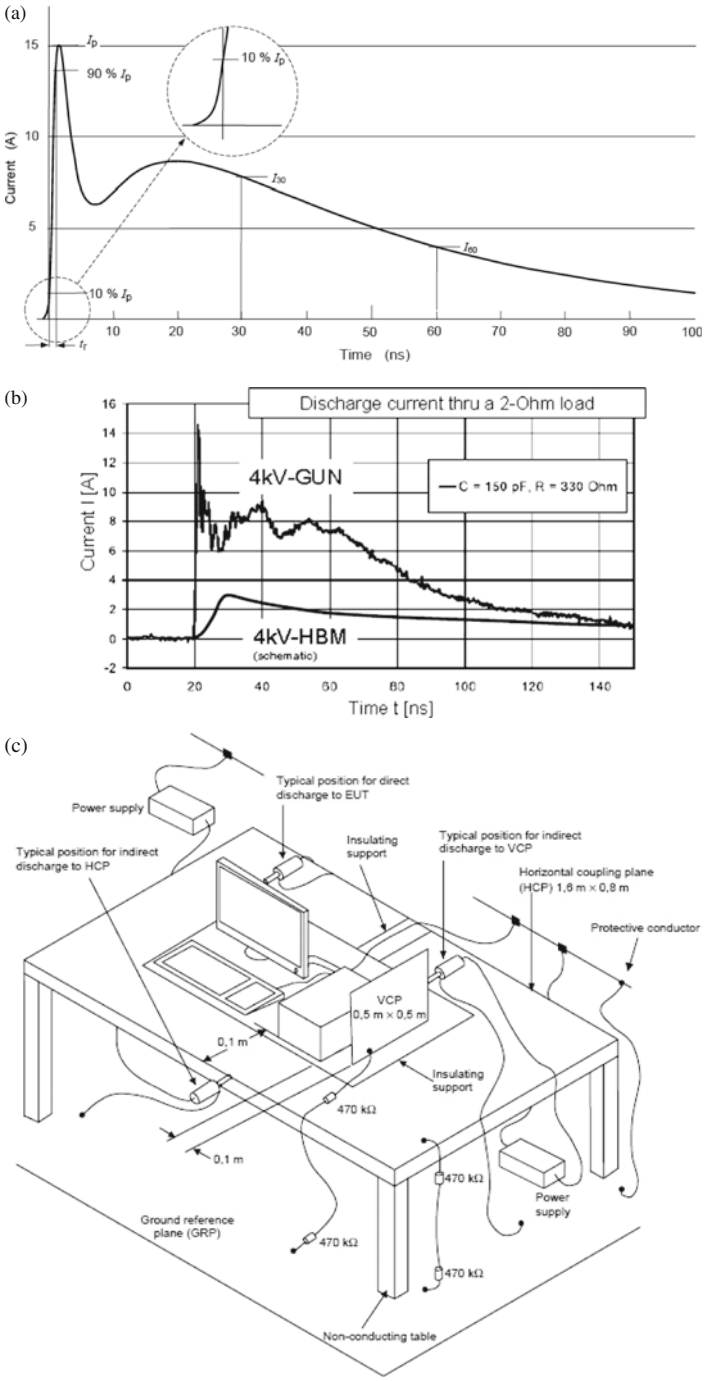


Fig. 8.4 Typical output waveform of ESD simulator (4 kV) (**a**) experimental waveforms for IEC61000-4-2 and comparison with HBM (**b**) and setup for IEC61000-4-2 system-level test (**c**)

From a practical ESD design point of view, a larger ESD current level is not a major problem in the case of system-level ESD stress. It can be handled by a corresponding width scaling of the ESD clamps to support system-level ESD current. However, the task that provides major challenges for on-chip system-level ESD design is withstanding ESD stress under power-on conditions.

While component tests are only performed with the IC unpowered, according to standards, the system test is done with both the unpowered system and with power applied to the functioning device. The subsequent requirements are that there must be no physical damage and the device must keep working normally after the stress has ended. This means latch-up must not occur on either the power or signal pins, as that will require system reset or power-down to bring the system back to normal operation.

Finally, unlike HBM testing, IEC stressing is done in two ways: contact discharge and air-gap (spark) discharge. Due to energy dissipation (loss) to the air, the discharge voltage requirement is higher for the air-gap test.

– Class 1 – 2 kV contact, 2 kV air discharge
– Class 2 – 4 kV contact, 4 kV air discharge
– Class 3 – 6 kV contact, 8 kV air discharge
– Class 4 – 8 kV contact, 15 kV air discharge

8.1.2.3 System-Level Test Setup

A calibrated waveform is usually realized by commercial testers. The test setup for a system test is defined by the standard document and is rather non-trivial. It includes the gun that provides discharge to the system, which is placed on the isolation mat on top of the table with the metal plate grounded through the 470 KΩ resistor (Fig. 8.4). It is clear that such a setup is based upon the coupling of the system with the ground plane. While the dielectric mat for the test is defined by the standard, the coupling will depend on system size and ground isolation. In particular, a portable system with a battery power supply might provide a totally different ESD current path than a similar system with a grid-connected power supply module that provides a low-impedance ground connection.

A comparison between the component and the system-level ESD stress is summarized in Table 8.1.

In each IC component used in the design of the system, there are different classes of pins that can be subdivided according to specific system-level ESD robustness (Fig. 8.5). The most common pins are intended for the on-board inter-chip connection. These pins are not critical for system ESD robustness and only need to meet the component ESD requirements. The second group of pins is attached to the signal busses connecting several PCBs. These pins are potentially exposed during repair and maintenance, which can be done violating EPA requirements.

Table 8.1 Comparison of device and system-level ESD stress

Function	Device level ESD test	System level ESD Test (IEC)
Stressed pin group	All pin combinations	Few special pins
Supply	Unpowered	Powered & unpowered
Test methodology	Standardized	Application specific
Test set-up	Commerical tester & sockets	Application specific
Typical qualification goal	1 ...2 kv JEDEC HBM	8 kv Contact (IEC 61000-4-2) 15kv Air (IEC 61000-4-2)
Corresponding peak current	0.65 ... 1.3 A	> 20 A
Fallure signature	Destructive	Functional or destructive

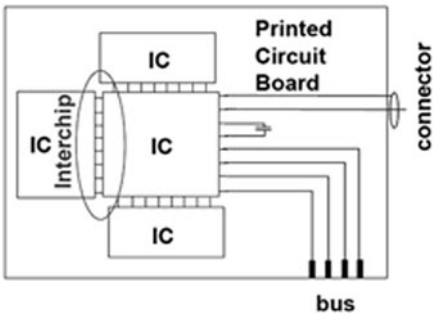


Fig. 8.5 Illustration for the three groups of pins with different system-level ESD requirements

The third group of pins is the most critical for system-level ESD; these are the pins directly or indirectly attached to the external connectors, for example, exposed USB pins. Depending on other connected components, these pins will experience either the whole or a partial level of system-level ESD stress. Typically, for such a group of pins, some increased ESD protection level is required.

8.1.2.4 Cable Discharge Event (CDE)

Cable discharge events occur during operation in the ports of systems, both in powered and in non-powered conditions. In real life, conditions for the primary charge can usually be formed due to the triboelectric charging of the cables handled across the facility prior to installation or during transportation from the manufacturer.

The standards for CDE are not yet established and therefore specific custom requirements, involving the twisted cable pair precharge level and the length of the cable (for example 2 kV and 200 m, respectively), are often applied to define the

test setup. The CDE is usually composed of two-stage events. The first part replicates the connection of the charged cable to the pin and the second part replicates the creation of short-circuit conditions on the opposite end of the connected cable.

When charged cables are connected to electronic products (or vice versa), the discharge is observed through the relatively low impedance of the cables. Cable discharge events (CDE) generally have similar waveforms for the initial discharge to the IEC, which represents the fast discharge of the low impedance part of the cable closest to the pin, and the discharge of the remaining part of the cable with a larger delay and impedance. Due to reflections from the unterminated cable ends, the follow-up current from the energy stored in the cable is in an oscillatory waveform, the frequency of which depends on the cable length. Today, ESDA WG14 is tasked with developing a standard practice for CDE, followed by draft and review.

An additional comparison of different standard pulses and the system tests are presented in Table 8.2. In particular for the presented data, it is clear that different system-level electrical overstress requirements may greatly exceed the accumulated charge level of the component-level specification by several orders of magnitude.

8.1.3 Transient Latch-up During a System-Level Event

One of the typical examples of different solutions for the IEEE 1284 interface standard is presented in Fig. 8.6. The original solution requires an external transient voltage suppressors (TVS) discrete ESD protection component for interface reliability. Plugging and unplugging printer cables can subject the printer ports to high electrostatic discharge conditions, which may ultimately render the printer inoperable.

External protection diodes add cost and consume PCB space, increasing the size of system blocks and eventually the cost of the system itself. Therefore, a more advanced solution represents system-level ESD protection integrated into the chip (Fig. 8.6).

Thus, introduction of system-level protection for the specified pins is an admirable goal. However, a new challenge in meeting the test specifications for a powered system arises due to the transient latch-up induced by the ESD devices and other parasitic on-chip components, as a result of the high current injected into the components.

As can be seen in the previous chapters of this book, in the case of component ESD protection requirements, local ESD protection can often be successfully realized using snapback clamps. These clamps simply turn off at the end of the ESD discharge event. This allows application of small footprint ESD protection clamps with holding voltages below the power supply level, or voltage levels at the pin in operation conditions.

However, in the case of power-on conditions, self turn-off may not occur. Self-turn-off will generally be a function of the load resistance at the pin, the difference between the minimum holding voltage and power supply level, and the critical turn-off current of the clamp.

Table 8.2 Comparison of device and system-level ESD stress for different pulses [141]

Standard/Pulse	Application	V[V]	Duration (10–90%)	# Pulses	Rl [Ω]	C [pF]	Ipeak [A]	Charge	Charge rel. to HBM
RC discharge generated pulses									
"HBM" JESD22-114	Component	8 k	150 ns	1	1 5 k	100	5.3	800 nC	1
"MM" JESD22-116	Component	400	16 MHZ	1	–		>8	0.8 nC	1 0E−03
"CDM" JESD22-101	Component	75	0.8 ns	1	1		8.5–17	–	–
"GUN" IEC 61000-4-2	System	8 k	120 ns	10	330	150	33	1.2 µC	1.5
ISO/TR 10605 inside	System/Vehicle	8 k	1µs	3	2 k	330	30	2 64 µC	3.3
ISO/TR 10605 outside	System/Vehicle	8 k	360 ns	3	2 k	150	30	1.2 µC	1.5
Voltage generated pulses									
ISO 7637: 1	System	−100	2 ms	5000	10	–	10	20 mC	2.5E + 04
ISO 7637: 2	System	100	50 µs	5000	10	–	10	0.5 mC	6.3E + 02
ISO 7637: 3a	System	150	100 ns	3.6E6 (lh)	50	–	3	300 nC	0.375
ISO 7637: 3b	System	150	100 ns	3.6E6 (lh)	50	–	2	200 nC	0.25
ISO 7637: 5 (load dump)	System	87	40–400 ms	1.10	2	–	43	17 C	2 lE+06

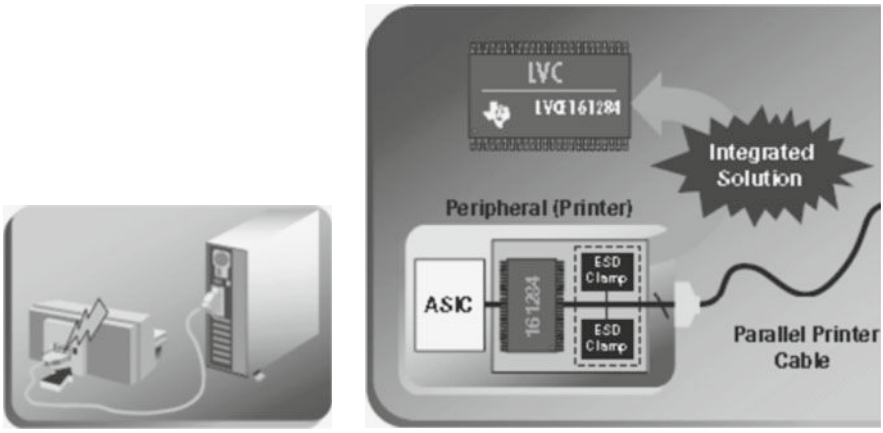


Fig. 8.6 Integration of the external protection diodes into the chip to provide protection in accordance with IEC 61000-4-2 on the connector interface pins

Some of these parameters are rather hard to measure. The minimum holding voltage measured using a 50 Ω TLP system may not correspond to the real value at a higher load resistance. Short pulse measurements of the turn-off current may not correlate with the actual current at long pulse conditions, where current filamentation phenomena may take over the current redistribution effect.

Thus, product verification of the on-chip system solutions is rather critical. In practice, the ESD protection cell is expected to provide an appropriate system-level current capability with the holding voltage of the used ESD device above the power supply level. In the opposite case, depending on the load resistance of the pin, a transient latch-up may occur. In this case, after an ESD event, the current through the snapback device in the high-current state will be continuously supported by the power supply, keeping the ESD device in high-conductivity state. The following shunting of the pin may result in loss of functionality or heat generation, and the high current density may result in accelerated degradation and permanent burnout of the ESD device in the latch-up state.

The waveforms for such an event are illustrated in Fig. 8.7 for the latch-up event induced in a parasitic SCR structure formed by a guard ring structure connected to VDD and V_{SS} busses [140]. At the pulse amplitude of up to 18 V, the device always returns to the original low current state. However, if the voltage is increased to 20 V and then reduced down to a lower level of ~3 V, the latch-up between VDD and V_{SS} lines is observed, due to high-current path formed.

Additionally, a numerical simulation for devices capable of hot plug-in (discussed in Section 8.3) also illustrates the transient latch-up effect.

An additional important comment should be made: in real application, many systems may be exposed to other kinds of electrical overstress with significantly higher power than ESD pulses.

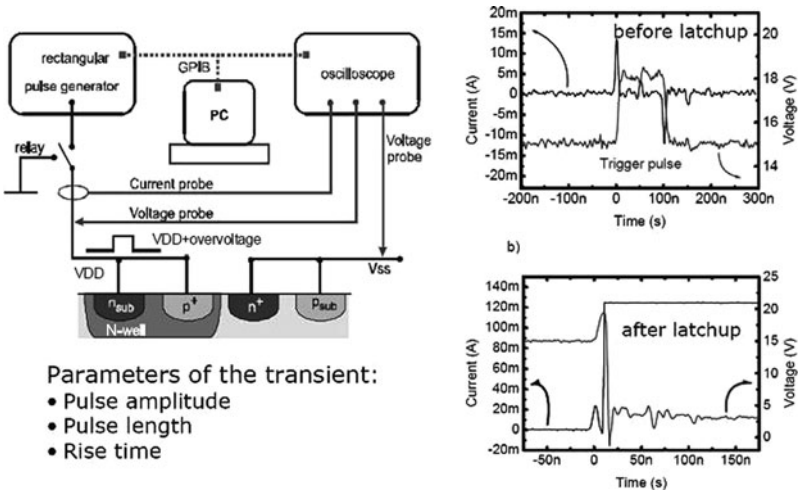


Fig. 8.7 Illustration of the transient latch-up in the chip with a parasitic p–n–p–n structure

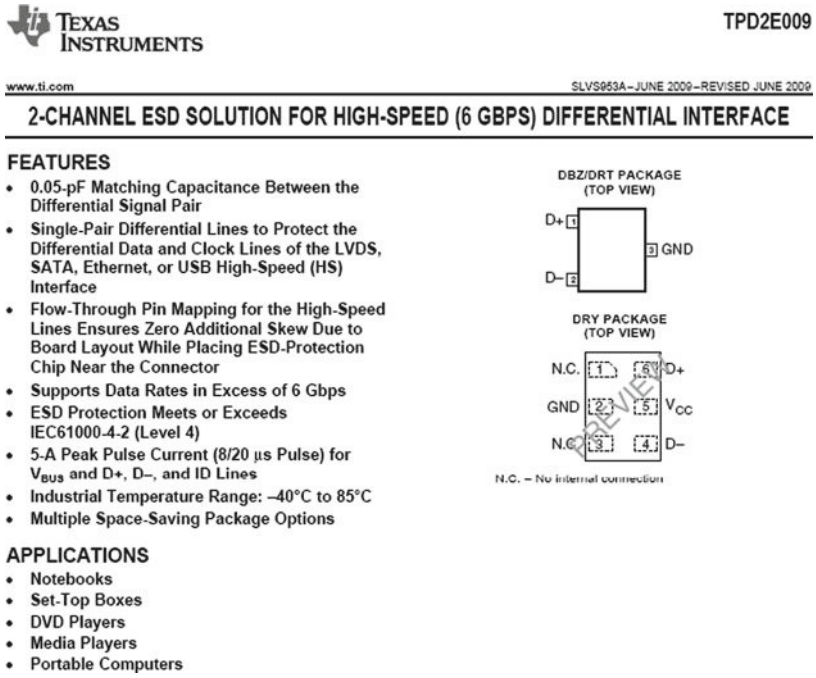
ESD structures (both on-board and on-chip) are first hit by these pulses. In general, these devices are not designed to withstand arbitrary overstress pulses. Therefore, robustness against EOS pulses has to be specified and taken into account explicitly in IC and board design.

8.1.4 System-Level Protection Components

To help system-level design, there is a great variety of transient voltage suppressors (TVS) available on the market. These suppressors essentially represent a discrete version of the ESD protection clamp with parameters defined by the datasheet. In the recent years, more sophisticated modules that combine arrays of suppressors, targeting specific automotive, high speed, low parasitic, integrated EMI filters, and other requirements became available. An example data sheet for such a component is presented in Fig. 8.8.

The TVS are used on PCB to provide a primary current path for the system terminal connected to the IC system pins, and often include additional separation components to realize a two-stage ESD protection network.

One of the alternative ways to manufacture suppressors in silicon process technology is based upon polymeric material with nanoparticles that provide extremely low parasitic capacitance, in comparison with semiconductor-junction base suppressors [142].



TEXAS INSTRUMENTS

TPD2E009

www.ti.com

SLVS963A–JUNE 2009–REVISED JUNE 2009

2-CHANNEL ESD SOLUTION FOR HIGH-SPEED (6 GBPS) DIFFERENTIAL INTERFACE

FEATURES

- 0.05-pF Matching Capacitance Between the Differential Signal Pair
- Single-Pair Differential Lines to Protect the Differential Data and Clock Lines of the LVDS, SATA, Ethernet, or USB High-Speed (HS) Interface
- Flow-Through Pin Mapping for the High-Speed Lines Ensures Zero Additional Skew Due to Board Layout While Placing ESD-Protection Chip Near the Connector
- Supports Data Rates in Excess of 6 Gbps
- ESD Protection Meets or Exceeds IEC61000-4-2 (Level 4)
- 5-A Peak Pulse Current (8/20 μs Pulse) for V_{BUS} and D+, D–, and ID Lines
- Industrial Temperature Range: –40°C to 85°C
- Multiple Space-Saving Package Options

APPLICATIONS

- Notebooks
- Set-Top Boxes
- DVD Players
- Media Players
- Portable Computers

DBZ/DRT PACKAGE
(TOP VIEW)

D+ 1
D– 2
3 GND

DRY PACKAGE
(TOP VIEW)

N.C. 1
GND 2
N.C 3
6 D+
5 V_{CC}
4 D–

N.C. – No internal connection

Fig. 8.8 Example of the datasheets for transient voltage suppressor

8.2 On-Wafer Human Metal Model Measurements

The material of this chapter is focused specifically on the on-chip system-level protection, as it is a major trend in portable system design. On-chip system-level protection creates advantages in both cost and appeal of small form factor products.

At present, characterization methods for system-level and component-level ESD stresses exist in parallel, without a real correlation between the testing results. In [143, 144], the correlation between CDM, HBM, and electrical overstress testing at the component level and tests that use different discharge networks at the system level has been studied. A correlation between system-level and component-level ESD characterization has been found. However, these results are obtained for only one type of technology and cannot be generalized for other applications.

When system-level pins are identified, corresponding design actions are taken to create an ESD cell with an adequate width toward the appropriate current level. The cell must not only target the appropriate ESD protection window, but must also meet the size and holding voltage requirements. Such design activity involves a significant amount of test chip-level experimentation and comparative analysis of the stand-alone cells. It would be rather inefficient to conduct such analysis on the IC or, especially, the system level.

At the same time, testing of the individual semiconductor components is essentially outside the scope of the IEC61000-4-2 standard. One of the typical decisions applied by manufacturers in overcoming this issue during component development is the fabrication of IC's with an elevated component ESD protection requirements for system-level interacting pins, for example, 4, 8, or 15 kV HBM.

However, as will be shown in this section, the correlation between component and system-level ESD pulses is not a given.

Another alternative is to test packaged devices using system-level tools. In this case, perhaps only the worst-case scenario can be reproduced for the component pin discharged to the ground, since the capacitive coupling of the system through the dielectric mat is hard to replicate.

Therefore, the human metal model (HMM) alternative has been proposed to create a recommended practice of testing semiconductor components and devices using the IEC (EN) 61000-4-2 basic. To allow the IC manufacturer to predict the ESD performance of their products under system-level stress conditions, the ESD Association (ESDA) has proposed a new measurement method: the HMM [145, 146, 147, 148]. The objective of the HMM initiative is to establish a common reproducible basis for evaluating the performance of electrical and electronic equipment when subjected to electrostatic discharges. In this section, HMM tester for on-wafer measurements is described.

8.2.1 On-Wafer HMM Tester and Equivalent Circuit of the Pulse

HMM reproduces the ESD discharge caused by a human touching a pin of a grounded electrical component with a metal tool.

The goal of this new HMM standard practice (SP) document is to simulate a system-level type of ESD stress at the component level. The focus of this SP is on components that have a direct connection to the stressed pin, such as interface circuits, external ports.

Studying the ESD performance of such components during the design phase requires on-wafer ESD measurements. Using an on-wafer HMM testing setup enables the study of protected ESD components under system-level stress conditions without prior packaging.

A version of the HMM ESD tester is presented in [144] using an on-wafer testing method. With the proposed setup, new results are obtained from the devices manufactured in both advanced CMOS and high-voltage technologies. The characterization results are further compared with those of the commonly used HBM testing.

HMM simulates a system-level type of ESD stress at the component level. Stress waveforms with a shape defined in the IEC 61000-4-2 standard document are applied to the device under test (DUT). The HMM standard practice, currently developed by the ESDA standards committee WG 5.6, uses an ESD gun as the pulse generator. Such an ESD gun contains a discharge circuit that conforms to the IEC61000-4-2 system-level ESD standard.

ESD guns have several disadvantages when used to test components. One issue is the electromagnetic field produced by an ESD gun during the application of a stress. The spread of the radiated field is not known and varies strongly between different gun models [145]. Also, the form factor of most of the available gun models is not suitable for the integration of such a gun into a component-level measurement setup.

An original on-wafer HMM testing setup using an ESD gun as the pulse source is presented in the standard (Fig. 8.9). Recently, this setup has been improved by replacing the ESD gun with a novel HMM testing module (Fig. 8.9b). This module, mounted on a standard wafer prober, delivers a discharge stress waveform according to the IEC61000-4-2 standard. Due to the specific hardware, electromagnetic disturbances do not occur during the discharge.

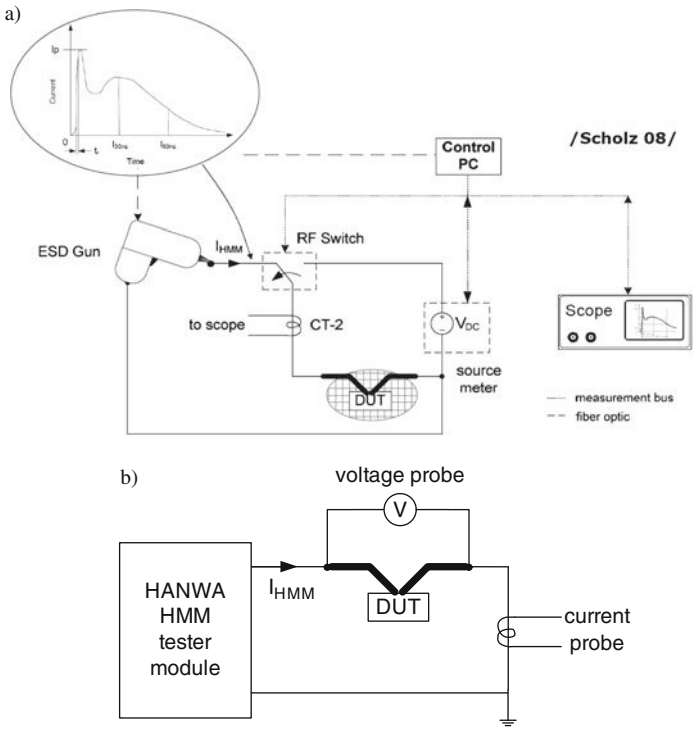


Fig. 8.9 Wafer-level HMM measurements setup with a commercially available gun tool (**a**) and specially developed HMM module (**b**) [143]

8.2.1.1 Equivalent Circuit for HMM Simulation

With this improvement and the addition of voltage and current probes, it is possible to simultaneously measure voltage and current waveforms during an HMM stress. The current is measured as a function of time using a Tektronix CT-2 current probe (Fig. 8.9). In [149], the current probe is capable of accurately measuring the HMM

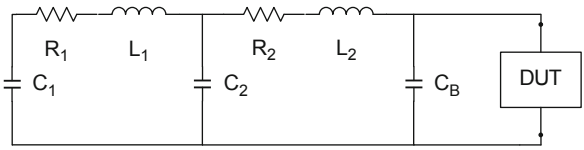


Fig. 8.10 Equivalent circuit for HMM pulse with two precharged capacitors C1 and C2

current waveform up to the required 30 A of peak current for a HMM precharge voltage of 8000 V.

The equivalent circuit for the HMM pulse is presented in Fig. 8.10, with parameters of the components summarized in Table 8.3. The ESD pulse equivalent circuit contains two precharged capacitors: C1 with a full precharge and C2 with 60% of the precharge level from the original C1 precharge.

Table 8.3 Parameters of the equivalent circuit for HMM pulse circuits (Fig. 8.10)

R1	210 Ω	Defines initial current peak
C1	120 pF	Capacitance with full precharge
L1	3 μH	Defines quasi-static current level
R2	120 Ω	Defines quasi-static current level
C2	30 pF	Capacitance with 60% level of original precharge
L2	100 nH	Defines rise time
CB	30 pF	Board capacitance

Such an equivalent circuit adequately correlates with both the specified standard parameters for IEC61000-4-2 waveforms mentioned above and the experimental data (Fig. 8.11).

8.2.2 HMM-HBM Component Correlation

Using the advanced wafer-level setup [143], typical ESD protection structures have been characterized in order to demonstrate a correlation between device responses to system-level ESD stress and to component-level ESD stress. For component-level stress, HBM pulse results have been taken for comparison. It has been found that the correlation factor strongly depends on the type of device.

The available commercial voltage probes are not capable of measuring high voltages with a high bandwidth. A new type of voltage probe was developed for this purpose. This voltage probe allows the measurement of an about 2000 V peak voltage with the bandwidth required for HMM measurements and enables the measurement of the turn-on behavior of an ESD protection device under HMM stress.

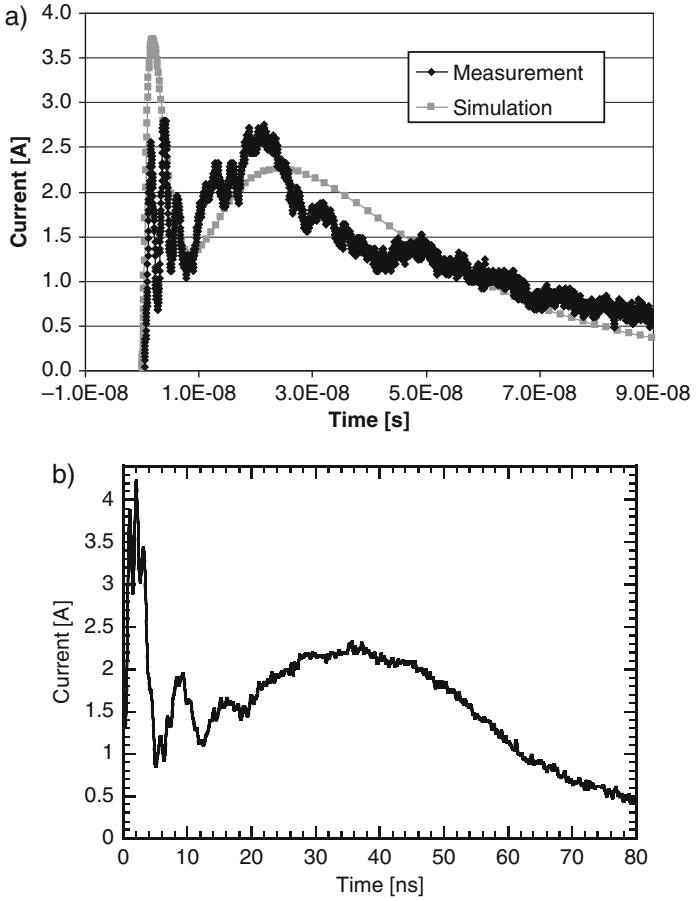


Fig. 8.11 HMM simulation for short-circuit conditions of 1 kV HMM pulse compared to measured HMM current waveform from the setup with commercial gun (Fig. 8.8a) (**a**) and advanced setup with Hanwa HMM module (Fig. 8.8b) (**b**)

Pass–fail measurements have been carried out on typical ESD protection structures, such as diodes, silicon-controlled rectifiers, and high-voltage ESD clamps, to study the capabilities of the HMM on-wafer measurement setup. To enable a comparison of the various methods, fresh samples of the same device have been characterized with the commercial HBM on-wafer HANWA HED-W5000M tester.

Only the waveforms of the maximum precharge level before device failure during HBM and HMM stress are compared below. These levels are equal to the maximum stress a device can withstand without being damaged or becoming nonfunctional. Thus, this methodology guarantees the comparability of measurement results when the correlation between the stress levels of two ESD testing methods is unknown.

To obtain the correlation factor, the precharge voltages obtained at device failure during HBM and HMM testing are divided. Voltage waveforms have been captured to study the transient device behavior under different stress conditions. To remove the additional voltage drops due to parasitic elements in measurement setups, calibration methodology [144] is applied to the on-wafer HMM setup.

8.2.2.1 Diodes Under HMM Stress

For diodes, the current waveform and DC characteristic are taken from devices manufactured in a 90 nm CMOS process technology.The diodes have a width of 50 μm with a diffusion isolation space of 0.5 μm.

The functional leakage test between the HMM pulses of elevated amplitude and corresponding current waveform at the step before failure is presented in Fig. 8.12. The results demonstrate a correlation factor of 3.7 between HBM and HMM testing. In HMM, the device can withstand a maximum current of about 7.8 A, while the maximum current in HBM is about 4.9 A (Fig. 8.12b).

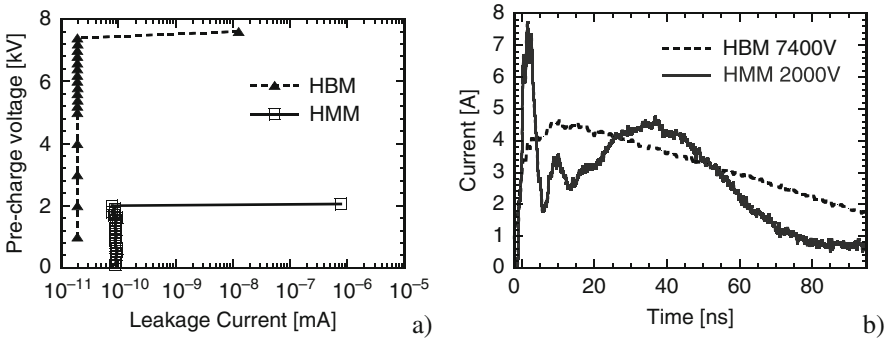


Fig. 8.12 Leakage evolution at negative 0.2 V bias during HBM and HMM tests (**a**) and overlay of HBM and HMM current waveforms obtained from a diode right before device failure (**b**)

8.2.2.2 LVTSCR

A different correlation has been obtained from test results of LVTSCR devices manufactured in the same 90 nm CMOS process. The devices have a width of 50 μm and a gate length L=0.25 μm. In this case, a correlation factor of ∼5 between HBM and HMM testing has been measured.

The overlay of HBM and HMM current waveforms just prior to device failure (Fig. 8.13b) shows that a particular device can withstand a maximum current of only ∼3 A in both cases. Unlike for the diode, the maximum thermal stress during HBM and HMM is different for the LVTSCR. The device is less robust under HMM stress due to the much faster rise time of the HMM pulse, which causes higher overshoots to occur during device turn-on. One of the possible explanations for this effect is

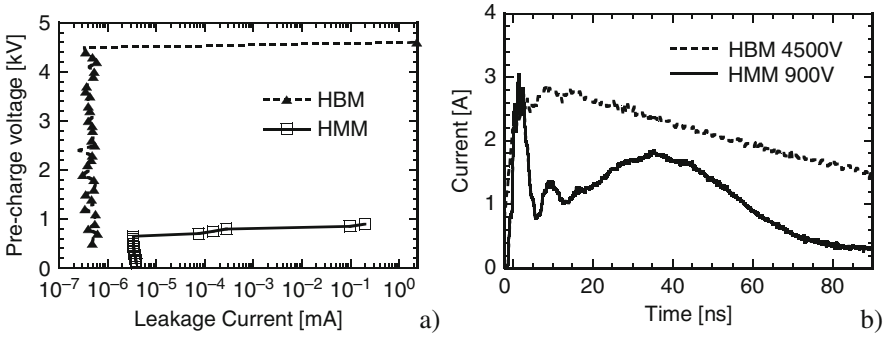


Fig. 8.13 Leakage evolution at 0.6 V bias during HBM and HMM testing (**a**) and overlay of HBM and HMM current (**b**) in LVTSCR devices

that the higher overshoots due to the first peak cause a failure in the device, which is already at lower current levels in HMM than in HBM.

8.2.2.3 High-Voltage ESD Clamps

A different correlation factor has been observed in high-voltage ESD clamps manufactured in a 50 V BCD technology and stressed according to the same methodology as low-voltage devices.

The measured correlation factor between HBM and HMM is about 1.5. This factor is significantly different from that of the measured low-voltage CMOS diodes. A comparison of the current waveforms captured just prior to device failure (Fig. 8.14b) reveals more differences. During the first 70 ns, the HMM current is significantly higher than the HBM current. After 70 ns, the HMM current falls below the level of the HBM current. In HMM, the device can withstand a maximum current of more than 6 A (initial peak), while the maximum current in HBM is only about 1.7 A, resulting in a higher robustness during HMM testing.

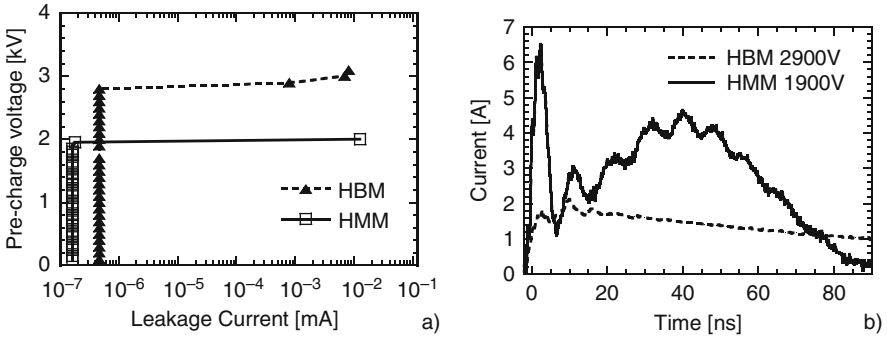


Fig. 8.14 Leakage evolution at 10 V bias during HBM and HMM testing (**a**) and overlay of HBM and HMM current (**b**) obtained from a lateral PNP device at the step prior to failure

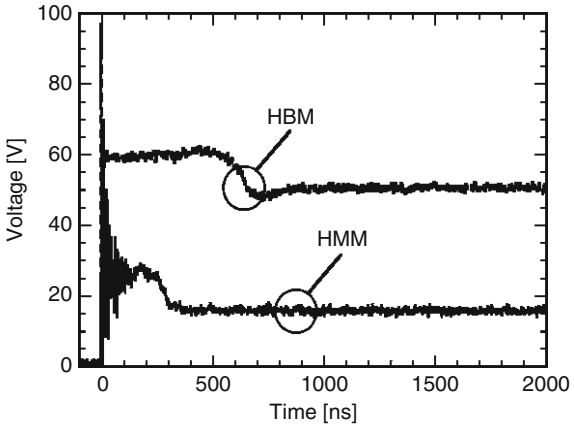


Fig. 8.15 Overlay of HBM and HMM current waveform obtained from a lateral PNP device right before device failure, stress level: HBM: 2500 V, HMM: 1800 V

This significant difference is due to the different behaviour of the device under different stress conditions. Figure 8.15 shows the voltage waveforms in the device during HBM and HMM testing. After turn-on, the voltage drop on the device is about 25 V during HMM stress and about 60 V during HBM stress, indicating different conduction mechanisms during both stress types. This creates less generated heat during HMM stress, thus making the device more robust, in comparison with HBM.

Thus, the correlation between system and component-level pulses is not universal and depends on device type and process technology. However, on-wafer HMM measurement setup can provide rather reliable information for the expected performance of system-level on-chip protection by recreating a similar stress environment to one that would be tested with an ESD gun as the pulse source.

8.3 On-Chip Design for System-Level Pins

8.3.1 Examples of Circuits with System-Level Protection

Historically, the first, most common products with system pins are products with the RS-232 interface standard. This standard has been designed to handle communication between two devices for short-distance (max 50–100 ft). Other examples include dual transceivers released to meet class 4 EN-61000-4-2; miniature electret microphone; high-speed bi-directional level shifters, ultra-low dropout voltage regulators; power management units, and display drivers, numerous automotive [153] and medical applications [154].

An example of the system-level ESD protection solution with EMI suppression is presented in Fig. 8.16. The circuit includes high-level ESD protection accomplished

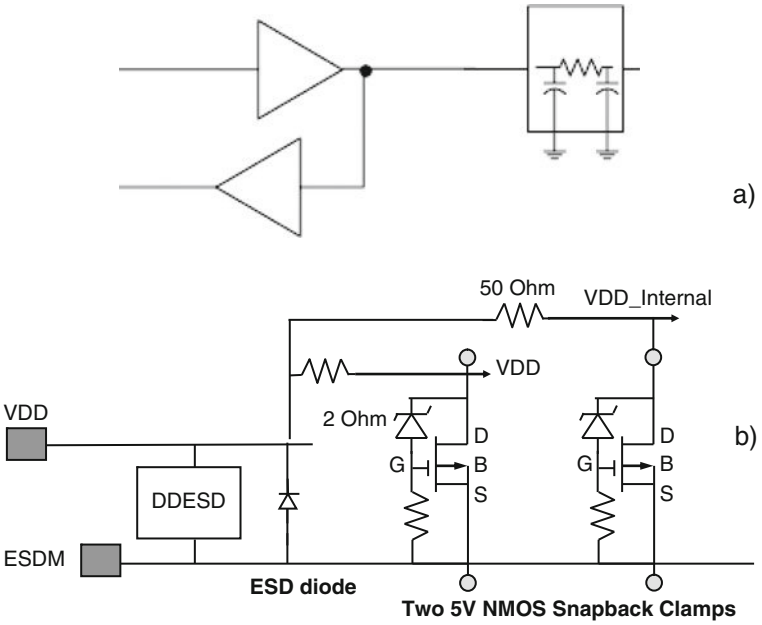


Fig. 8.16 EMI filtering with system-level ESD protection. Block diagram for EMI filter (**a**) and ESD protection network with dual-direction device DDESD (**b**)

with a dual-direction ESD device. The dual-direction functionality is required to avoid generation of EMI harmonics. After the EMI filter components, the second stage of ESD protection includes a 50-Ω separation resistor and an NMOS snapback clamp.

Another example of on-chip system-level protection requirement is for the input/output pins to be tolerant of a much higher voltage than what is usually applied to circuits fabricated in a low-voltage BiCMOS process. These cases can be found in a variety of analog applications for displays, automotive, and communications products.

In this case, it is desirable to provide ESD protection for circuits with a high-voltage tolerance in a low-voltage BiCMOS process without adding additional process steps to avoid increasing the cost.

One example is presented in Fig. 8.17. The microchip is fabricated in a 5 V BiCMOS process. The chip is intended to calculate the current through the inductive load by measuring a small differential voltage across the shunt resistor R_S. The common-mode voltage on R_S can be very high when the power FET is on. The voltage drop across the input resistors R_{I1} and R_{I2} shields the 5 V active components in the differential amplifier (DA) from this high-input, common-mode voltage. When the power FET is off, the current through the load returns through the freewheeling diode (D). This results in a negative-input, common-mode voltage on the shunt

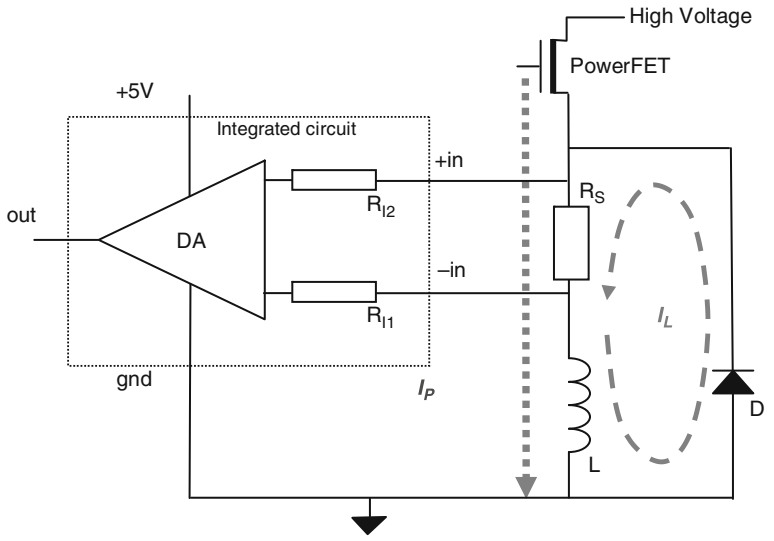


Fig. 8.17 Current measurement circuit in the application with an inductive load using a differential amplifier (DA) and on-chip resistors R_{I1} and R_{I2}

resistor. As a result, the full, required voltage range for the input pins is –5 V to +60 V.

Presence of the body diode inside typical BiCMOS ESD protection devices NPN BJT, BSCR, NMOS, LVTSCR, or active clamps with grounded P-substrate prevents proper chip performance. For example, if the circuit (Fig. 8.17) the current in the inductance (motor) circuit is measured by sensing a small differential voltage on the small shunt resistor R_S. High current through the body diode will create a large voltage drop on the circuit interconnects, making accurate circuit operation impossible. As a result, one of the basic devices for dual-direction operation should be similar to a diode AC switch (DIAC) [14]. The BiCMOS version of this dual-direction ESD protection device has been proposed in [150] and recently studied further in [42, 151, 152].

To improve voltage tolerance of ESD devices beyond process capabilities, a new method for increasing the breakdown voltage is proposed in [151], and is motivated by the goal of keeping the ESD devices as free structures, i.e. structures that can be realized using only layout changes. This method is based on the use of a "spotted" NISO blocking mask to create a corresponding spotted implant pattern. It is important that the NISO region is formed early in the process, before the NEPI deposition, and followed by the high-temperature anneals.

The total anneal produces an NISO diffusion profile of ~7 μm. As a result, even for the case of a spotted implant pattern (Fig. 8.18a, b) defined within the minimum dimension design rules (2 μm), the final doping profile is essentially quasi-uniform (Fig. 7.18c). This new, "diluted" NISO layer has an average doping that is lower

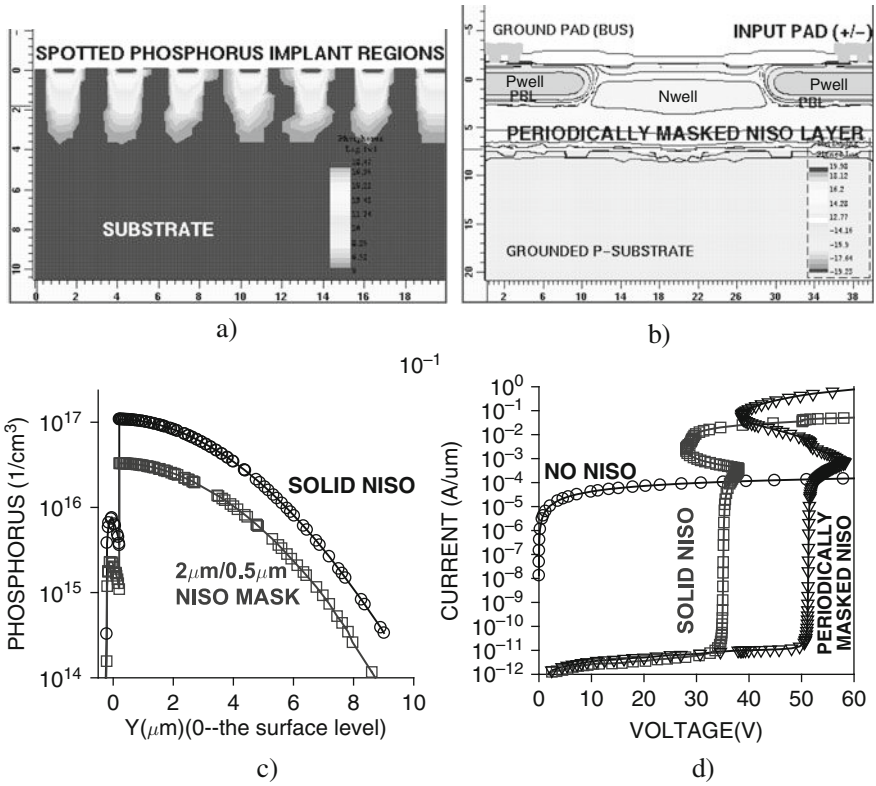


Fig. 8.18 The cross-section for the phosphorus profile after NISO implant (**a**), the final doping profile for the NISO region (**b**, **c**), and the calculated isothermal *I–V* characteristics (**d**) for a PNP triggered DIAC [42]

than the initial NISO implant (Fig. 8.18d) and can be used to form asymmetrical blocking junctions with higher breakdown voltages.

According to TCAD analysis, 50% blocking of the NISO implant by using the 2 μm × 2 μm mask creates a spotting pattern that produces approximately half of the NISO doping level, in comparison with the NISO region that is formed in standard processes (Fig. 8.18c). This can be used to substantially increase the breakdown voltage of the DIAC device to above 60 V in experimental design (Fig. 8.19).

If dual-direction voltage tolerance is required for the input pad, only the symmetrical device (Fig. 8.18b) can be modified to an asymmetrical bi-directional device (Fig. 8.20), saving additional space on the chip.

Thus, a voltage tolerance that significantly exceeds that of devices supported by the process technology can be achieved for system-level solutions.

A generic method for increasing the breakdown voltage consists of creating a diluted doping layer using a spotted implant mask, followed by anneals that result in a quasi-uniform doping layer formation. The diluted doping layer can also be

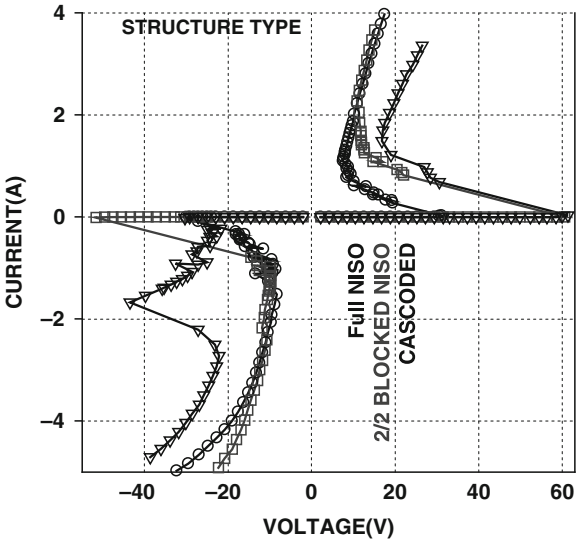


Fig. 8.19 Measured TLP characteristics of the reference devices with full NISO layer, diluted NISO layer, and the cascaded version

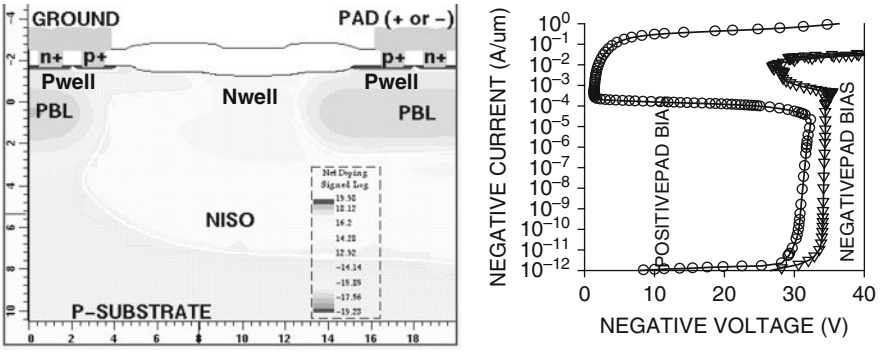


Fig. 8.20 The cross-section with the net doping profile and the calculated isothermal *I–V* characteristics of the asymmetrical dual-direction DIAC structure

used to obtain blocking junctions with a higher breakdown voltage than available in the BiCMOS process. The overall value of this technique is in providing a greater variety of ESD protection devices and leading to improved ESD protection window targeting for both BiCMOS and submicron CMOS processes with a triple well.

8.3.1.1 On-Chip and Suppressor ESD Device Interaction

The case of on-chip and discrete system-level protection interaction is illustrated in Fig. 8.21 [139]. An IC pin with normal ESD protection is protected by an additional PCB suppressor diode. Respectively, the majority of current during the ESD event is expected to flow through the PCB diode. The peculiarity of such an approach is related to the comparison of the characteristics of the PCB diode and the ESD protection on the chip. When the voltage drop on the PCB diode is lower than on the IC protection, the expected current path is valid. However, if the critical for failure voltage drop on the internal IC pin is lower than the level provided by the PCB diode, the internal ESD circuit can fail even if the PCB ESD diode itself can withstand ESD pulse.

Fig. 8.21c illustrates the possible effect of bus resistance, which may result in an unexpected burnout of the bus unless it is engineered to conduct system-level ESD current.

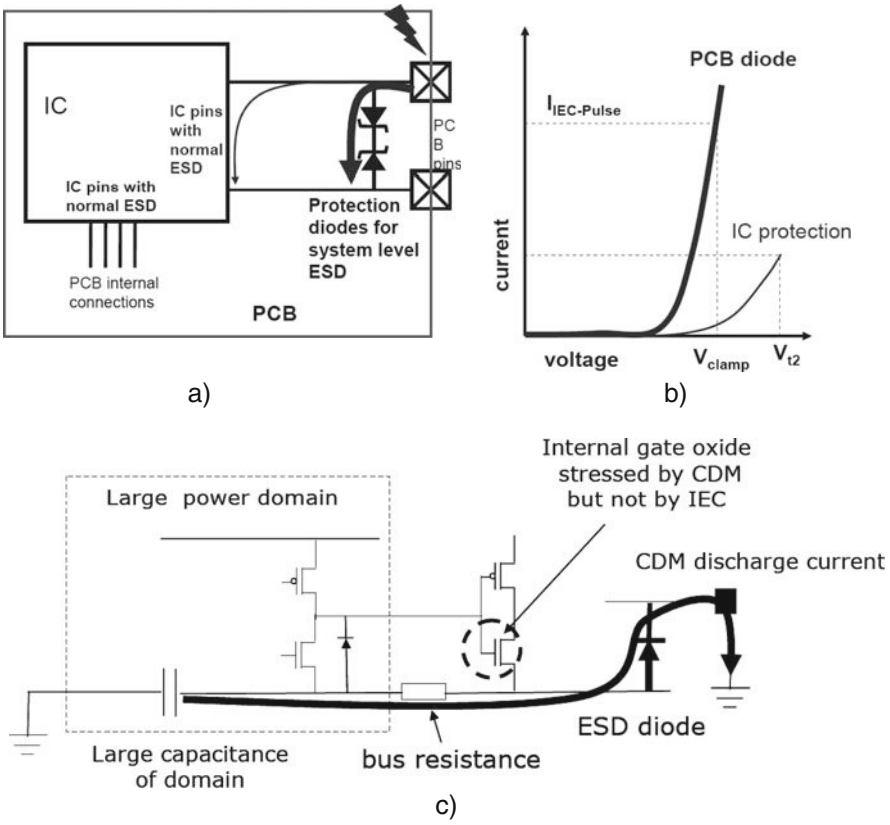


Fig. 8.21 Combination of the on-chip and system-level PCB protection [139]

With the on-chip system-level protection design, there is an area and performance trade-off. The performance of the end system is the real target. A learning cycle may be required to correctly target the "ESD window." Component and system tests must be carefully specified in advance. Additional external components may be required to prevent a latch-up event.

8.4 Hot Swap and Hot Plug-in

The problem of power sequencing and ground detection occurs in many systems. For example, mobile systems using a battery supply have a power up sequence that is limited by the input capacitance and the high resistance of the battery. Hot swap controllers provide limiting of inrush current and short-circuit protection during operation.

A particular aspect of system-level ESD protection problem is related to a hot swap and plug-in requirements. It is usually assumed that hot swap and plug-in events are performed in a condition that is at least partly EPA compliant.

The high holding voltage lateral PNP device has been previously experimentally discussed in Chapter 3 of this book. The disadvantage of such a device is its relatively large footprint.

As it has been mentioned above, the most reliable way to guarantee system stability against transient latch-up is to require a minimum on-state holding voltage above the power supply voltage level. This requirement also maintains system stability against short-term electrical overstress (EOS) events above absolute maximum voltage limits. Alternatively, the ESD protection clamp or network may be designed to guarantee both critical snapback and the minimum holding current levels above the current level that can be provided by the power supply at normal operation and test conditions. In this case the conditions for latch-up cannot be realized in principle.

A new compact PNP-SCR device-level solution utilizing a two-stage snapback approach is proposed in [49].

8.4.1 The Concept of Two-Stage SCR ESD Devices

In some hot plug-in and hot swap applications, as well as in some applications where system-level protection is required, the EOS and power supply current can be rather limited. Therefore, rather than define a minimum specified holding voltage for robustness against transient-induced latch-up, a minimum snapback current, for example, 1 A, may be chosen above the range of the power supply current. Below this current, the device will be robust against latch-up due to the high power required to induce snapback.

A consequence of the high holding voltage of a lateral PNP device is a required large footprint (1000 μm total width) due to its high on-state resistance in the conductivity modulation mode. This contrasts with the low-holding-voltage

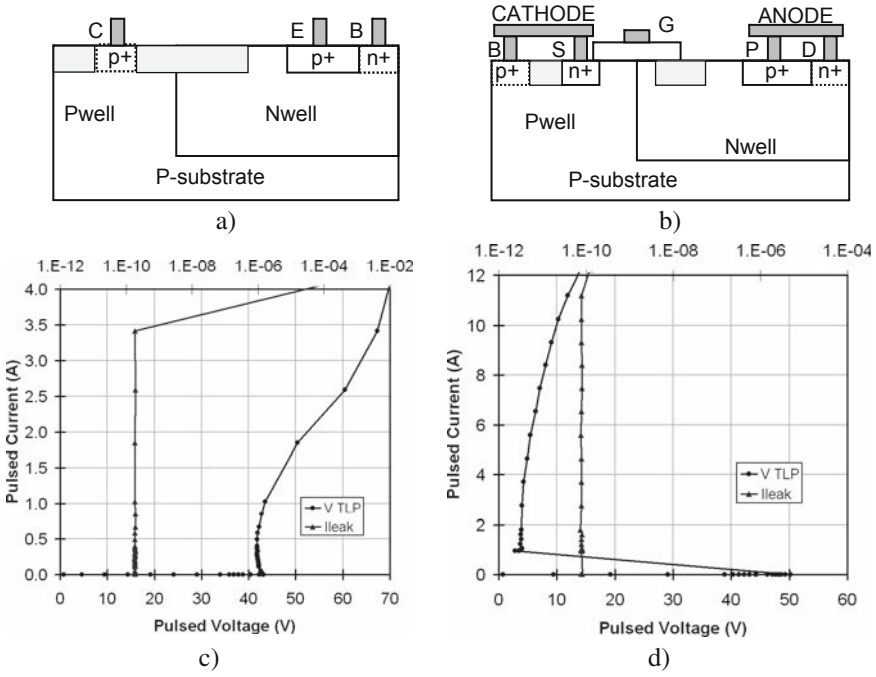


Fig. 8.22 Simplified cross-sections of lateral PNP (**a**) and DeMOS-SCR (**b**) devices, and the experimental TLP *I*–*V* characteristics (**c**) and (**d**), respectively

DeMOS-SCR device, which can be implemented in a smaller footprint. The lateral PNP also produces a bell-shaped voltage waveform during ESD stress that can challenge internal circuit protection.

Both the lateral PNP and DeMOS-SCR devices presented in Fig. 8.22 have beneficial aspects; therefore, it is desirable to design a two-stage device with the high holding voltage characteristic of a PNP at low current, and the low on-resistance characteristic of an SCR at high current.

To achieve such a two-stage operation in a practical manner, an additional isolation-limited n^+ emitter can be embedded in the p-collector region of the lateral PNP device. This can be achieved by modification of original device topology (Fig. 8.23a) by enclosing the additional n^+ emitter diffusion in the p^+ collector diffusion (Fig. 8.23b) or by creating interdigitated n^+ emitter regions (Fig. 8.23c). In both cases, the triggering current at which the device switches from PNP current modulation mode to SCR mode can be controlled by the n^+ emitter topology. In the topology of Fig. 5b, an increase in the enclosure parameter L_E should increase this triggering current. Similarly, in the topology of 5b, varying the number and relative lengths of the n^+ emitters will control the triggering current.

The cross-section of the device (Fig. 8.23a) is already shown in Fig. 8.22a. The cross-section for the topology (Fig. 8.23b) corresponds to the process simulation cross-section (Fig. 8.24a). The cross-section for the topology in Fig. 8.23c is roughly

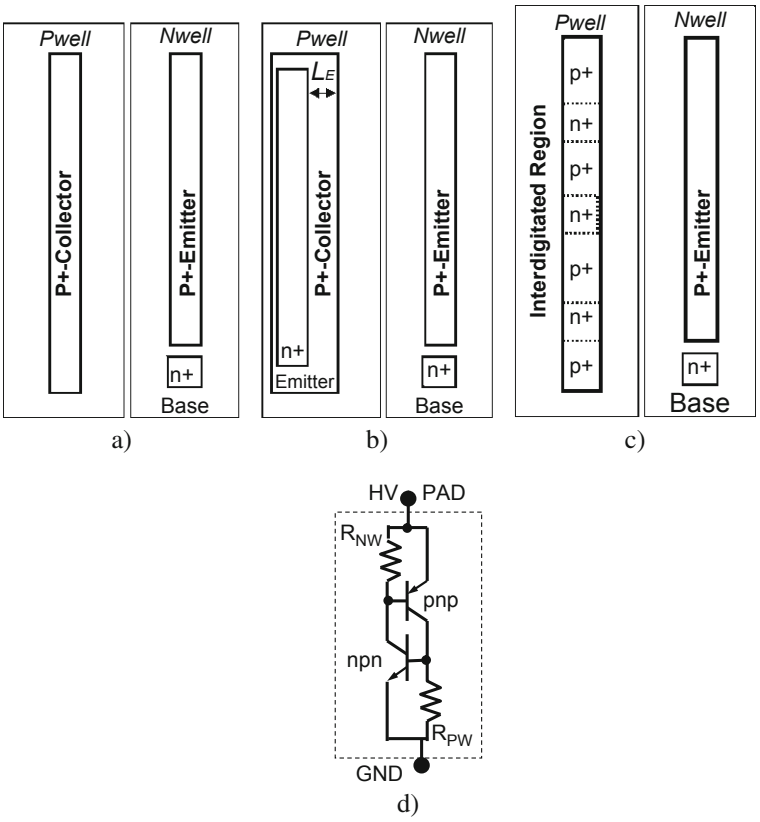


Fig. 8.23 Simplified original topology for the high-voltage lateral PNP ESD device (**a**) and implementation of the two-stage PNP-SCR devices using enclosed (**b**) and interdigitated (**c**) n^+-emitter; the equivalent circuit of the device (**d**)

a superposition of the two cross-sections referred to above, depending on the cutline position.

The equivalent circuit of the new device is still the same as for any SCR structure (Fig. 8.23d). For example, it is similar to the cross-section of the NLDMOS-SCR if the polygate region is excluded. However, the major difference is in the selection of the parasitic n–p–n and p–n–p structures that form the SCR and, in particular, the additional value of the two-stage snapback capability.

In the case of the NLDMOS-SCR, the triggering in the first stage is controlled by the parasitic n–p–n structure that has an already rather low holding voltage. At the given load characteristics during the device turn-on, the critical current is sufficient to initiate double-injection conductivity modulation from the n^+- and p^+-emitter junctions.

In the case of one of the topologies properly implemented for the two-stage SCR (Fig. 8.23a–c), the lateral p–n–p is the embedded component that first controls

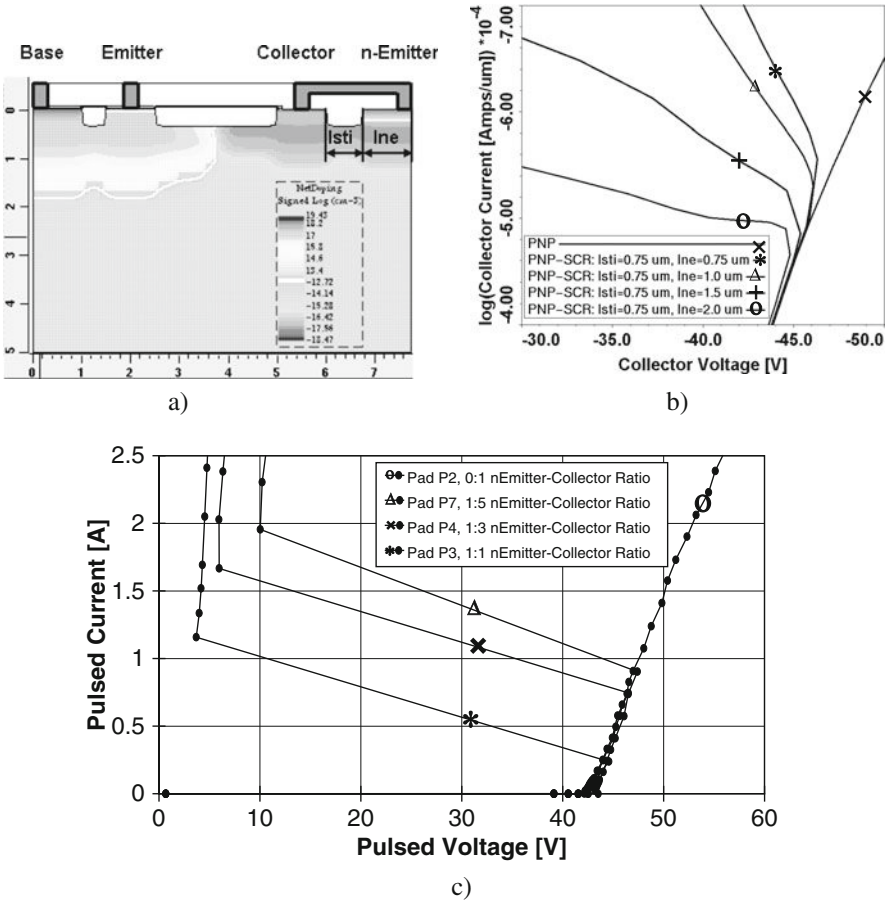


Fig. 8.24 Two-stage PNP-SCR device cross-section (**a**), with simulated *I–V* characteristics (**b**), and experimental dependence of TLP characteristics on n^+/p^+ interdigitation composition (**c**)

operation in avalanche–injection conditions. Until the current density in the structure becomes sufficient, the n^+-emitter junction is closed and device operation is similar to that of the lateral PNP device. The potential of the emitter junction rises over ~0.8 V and the injection of electrons becomes possible only at a certain critical current, followed by positive feedback for the double-injection conductivity modulation process from the p^+ and n^+ emitters similarly to the final state of the NLDMOS-SCR.

Thus, the sequence of the embedded parasitic pnp and npn components provides differentiation between the two-stage SCR proposed in this study and the known NLDMOS-SCR solution.

One of the expected tunable characteristics of the device is the critical triggering current in SCR mode. The higher the isolation of n^+-emitter, the higher the triggering current expected. The isolation level of the n^+-emitter is a layout-dependent figure of merit. For example, in layout topology (Fig. 8.23b), larger spacing L_E causes a higher isolation level. A similar statement is true for the reduction of the total amount of interdigitated minimum dimension n^+-emitter islands in the topology (Fig. 8.23c). Thus, an important degree of freedom is provided on the topology level to control the desired current and voltage before snapback into SCR conductivity modulation mode.

The suggested concept of the two-stage PNP-SCR clamp has been validated by physical process and device (TCAD) simulation (Fig. 8.24a, b), with successful experimental implementation in a 5 V 0.18 μm DGO CMOS process with 40 V extended drain devices (Fig. 8.24c).

A cross-section of the simulated device is presented in Fig. 8.24a. The structure is similar to the lateral PNP in the base and emitter region (at left), with the n^+ emitter embedded and isolated by the p^+ collector region (at right), to form an SCR device. The level of n^+ emitter isolation depends on the device topology and can be simulated. Greater n^+ emitter isolation results in a higher triggering current between the PNP and SCR stages (Fig. 8.24b).

In [49], experimental devices have been implemented using the interdigitated approach of Fig. 8.23c. The interdigitated regions of the n^+ emitter and p^+ collector form one stripe of the composite region. A decrease in the triggering current and holding voltage with an increasing n^+ collector to p^+ ratio observed in the TLP measurements of Fig. 8.23c is consistent with the simulated results of Fig. 8.24b

At the same time, all devices provided an on-state current suitable for system-level tests.

A mixed-mode simulation has been performed in order to demonstrate how the new solution will perform in power surge electrical overstress (EOS) events and in HBM operation. The output characteristics of the devices used in the simulation are presented in Fig. 8.25a. Two mixed-mode circuits used for validation are presented in Fig. 8.25b, c.

In the EOS event, a DC voltage source of VIN = 35 V is applied to the two-stage devices A and B with characteristics presented in Fig. 8.25d. The EOS pulse is produced by the current source I_{EOS} in a time domain close to fast ESD conditions. According to the transient characteristics with two-stage device "B" (with a low triggering current) remains in the low holding voltage latch-up state, while device "A" (with a much higher triggering current) returns to low leakage state (Fig. 8.25d).

The 8 kV HBM operation of both devices provides rather similar waveforms suitable for high-level ESD protection (Fig. 8.25e).

Thus, for hot swap and hot plug-in requirements, use of the two-stage HV PNP device with an embedded parasitic SCR structure is a rather promising solution. It enables design parameters that are controllable by layout for a two-stage operation in the PNP and SCR conductivity modulation modes.

Practical implementation of this new PNP-SCR clamp is achieved by simple topological control of the first-stage triggering current to a desired level.

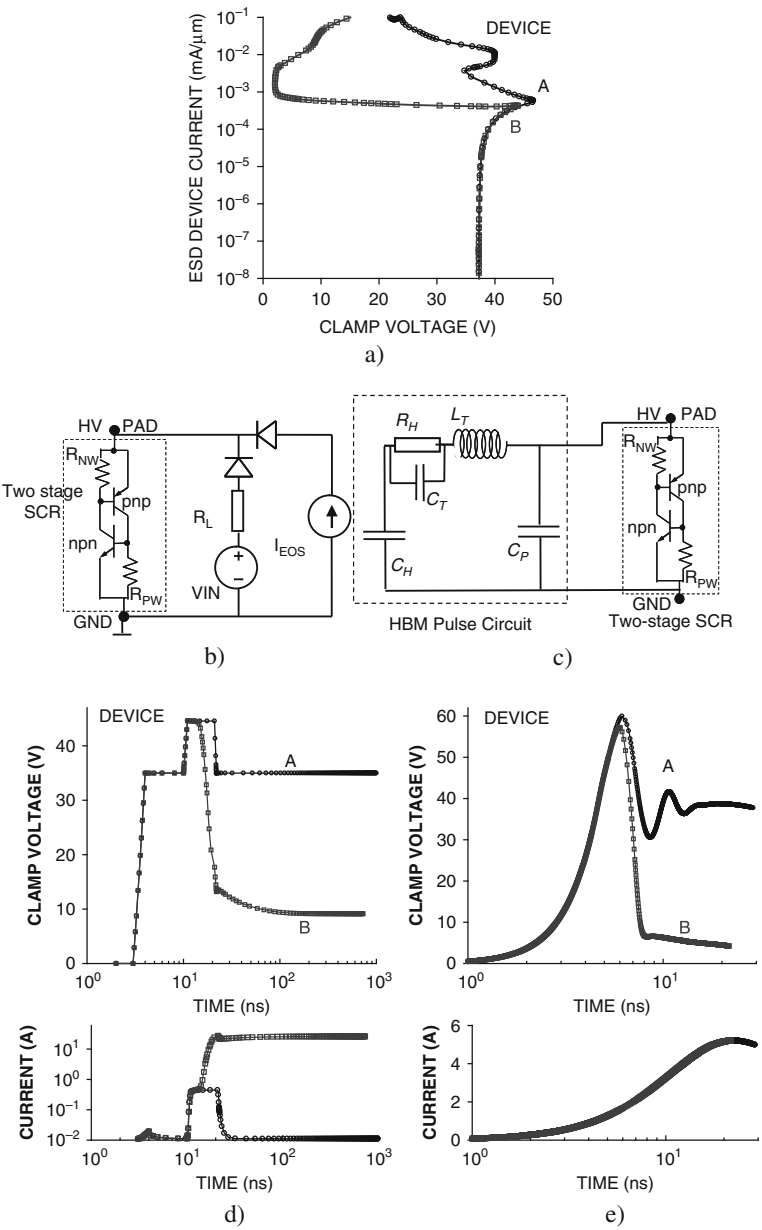


Fig. 8.25 Simulated DC *I–V* characteristics for two two-stage SCR's with different critical currents for SCR turn-on (**a**). Mixed-mode simulation circuits used to validate the hot plug-in solution in case of electrical overstress (**b**) and HBM ESD pulse (**c**); corresponding transient pulse waveforms (**d**) and (**e**)

In contrast to an SCR device, this solution is inherently insensitive to electrical overstress upto a desired level, as long as the overstress current is below the triggering current for transition into the SCR mode.

8.5 System-on-Package (SOP) Protection

To optimize the cost of the integrated components, a system-on-package (SOP) approach is used. With (SOP), the final packaged device is composed of two or more dies bonded in the same package. A typical example of the SOP is the integration of switching voltage regulators or power arrays in one package with digital blocks to create complex power management units (PMU's). A PMU essentially combines a co-processor unit made on expensive silicon with an inexpensive power module to support high-density digital design using 0.45–0.13 μm CMOS processes, for example, in the 0.5 μm BCD process. Similarly, because of simple cost considerations, the modules may include a combination of dies from high-voltage processes where the dense core logic devices are simply unavailable. A stacked die SOP example is presented in Fig. 8.26.

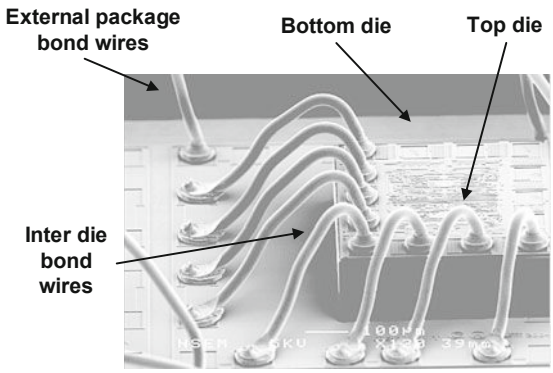


Fig. 8.26 Illustration of the shared interdie ESD protection

Since the ESD protection specification is defined for packaged parts, SOP design can be advantageous. First of all, the die pins not connected to the lead frame of the package do not need any protection at all.

An advantage can also be taken from the external pins of such an SOP. In particular, some of the more expensive external die pins can be protected either fully or at the first stage by the ESD network of the less expensive die.

Similarly, the least expensive silicon die can protect system-level pins, or additional dies with a TVS function can be introduced into the package.

8.6 ESD Robustness of Discrete Components

8.6.1 Discrete Components in High Reliability Systems

Electronic components form an integral and critical part of on-board systems and are present in vehicle controls, communications, navigation, radar systems, etc. Future aircraft systems, including electric aircrafts and Unmanned Aerial Vehicles, or the Next Generation Air Transportation Systems, will certainly rely more on electric and electronic subsystems and components. This new application will also result in the increase of the number of electronic faults with perhaps unanticipated fault modes. In addition, the move toward lead-free electronics and Micro-Electro-Mechanical Systems will further result in unanticipated anomalous behavior.

To improve the reliability of systems such as aircraft, assure adequate in-flight performance, and reduce maintenance costs, it is therefore imperative to provide system health awareness for the electronic components. To that end, an understanding of the behavior of deteriorated components is needed, as well as the capability to anticipate failures and predict the remaining life of the embedded electronics. The development and advancement of this capability is also relevant to multiple NASA exploration systems and long-endurance robotic space missions.

Generally, an understanding of the intrinsic and extrinsic degradation mechanisms of discrete devices is crucial for the adoption and application of health management to systems. Within the field of electronics, knowledge of semiconductor degradation under various system and environmental scenarios may be coupled with prognostic algorithms to predict the future state and time-to-failure of semiconductor components.

The existence of measurable extrinsic degradation precursors, pertaining to device packaging, has been well established in literature concerning power transistor devices. Intrinsic degradation precursors related to the physical properties of the semiconductor materials have also been observed. However, it is not widely known how degradation propagates as a function of the environmental conditions and various stressors, in particular, ESD stress. The attainment of such knowledge is critical for advancement in the field of power electronics health management and prognostics.

8.6.2 ESD Requirement for Discrete Components

ESD and EOS stress are major sources of damage and failure of discrete electronic devices, especially field effect transistors and IGBTs, which are extremely sensitive to such stress. In the case of low-voltage (<100 V) and low-power devices (<1 W), even package-level ESD events usually result in an immediate irreversible burnout. This event is easy to detect by a simple electrical parameter test. For high-power MOSFET and IGBT devices, the scenario is different. As shown below, at some pin

combination, these devices may indeed withstand not only the standard package-level stress but also the system level. However, even in this case, ESD robustness of the control electrode is negligible and requires an ESD protected area environment to guarantee the absence of handling damage.

Independent from the levels of ESD pulse that can cause local damage of the power device, another question is the final impact of the ESD events on the long-term reliability parameters of the stressed device. Depending on the nature of the induced defect, an accelerated degradation and corresponding reduction of the time before failure can be expected, due to the change of electric field distribution, current density, and lattice temperature in the location of the defect.

At the same time, ESD protection requirements are originally defined for integrated circuit components. A ESD specification of the most commercially available discrete components, for example, IGBT and MOSFET, cannot be found even for the control electrode.

Often, an intuitive assumption about ESD robustness of a discrete component is based upon the power and high-voltage parameters of the component in comparison with the ESD pulse. For example, with a component current level over 10 A, which significantly exceeds the package-level ESD current or the operational voltage of the component in a kilovolt range, the expectation is rather high.

To address the ESD robustness of a discrete component, test results for a discrete 600 V 16 A Insulated Gate Bipolar Transistor (IGBTs) device with an integrated flyback diode are presented in this section. It is demonstrated in [155] that ESD robustness depends on the terminal of this device. In the case of collector stress, the device can indeed easily outperform the ESD level up to a 16 kV ISO discharge. However, if the gate of the device is stressed, the device can only pass the 5 kV HBM test. Thus, the robustness of discrete power components is rather device and terminal specific.

It is demonstrated that ESD pulse stresses may create latent local damages that act similar to defects in the structure of the power device. Such stress-induced defects may not result in immediate parametric failure due to the change of the measured parameters from the datasheet range. While some deviation of the parameters can be observed within the datasheet limits, the long-term reliability of the component is significantly impacted. Thus, stress-induced defects do not result in formal parametric failures per datasheet specifications, but can be detected by the substantial change in the electrical characteristics when compared with the original device parameters measured prior to ESD stress.

8.6.2.1 ESD Effects on Power Transistors

Some results obtained by system-level stress tools are presented below for the commercially available 600 V 16 A IGBTs [155].

In power transistors such as power MOSFETs, the gate oxide is extremely susceptible to ESD events. Gate oxide stress above the critical voltage can cause structural damage to the gate oxide. This damage leads to leakage current levels

beyond the datasheet limits. Alternatively, it could also cause a moderate change in the device structure that may result in failure during future operation.

Thus, ESD damages are categorized as two types: (a) catastrophic and (b) latent. Devices with catastrophic damage are those that have been instantly destroyed at the time of the ESD event and therefore will not pass the functional tests that are part of the manufacturing or repair test process.

Latent damage occurs in devices when the ESD degrades the device, but not to the point of destruction. In this case, the changes in the power device structure are local and deviations of characteristics due to changes in the local region are hard to detect. Therefore, these devices could generally pass the functional tests, especially after only a single pulse ESD event. However, the reliability of components with latent damage is in question. One of the major resulting expectations is failure due to further ESD events or other forms of electrical overstress (EOS). Another possible impact is on the safe operating area, thereby reducing device robustness in standard application regimes.

The most common result of gate oxide overstress due to ESD events is the breakdown of gate dielectric material.

Usually, immediate failure does not occur until the gate-to-source voltage exceeds the rated maximum by an amount that may be in the range of two to three times the rated maximum.

During the manufacturing of the device, it can be assumed that structural defects are well controlled. Nonetheless, a small probability of a priori defect remains [156, 157, 158]. For discrete power device arrays, the detection of these defects is hard. An acceleration of the degradation processes can be expected for devices with prior defects that manifest themselves in an increased electric field, current density, and Joule heat generation. In contrast, defects that are induced during the operation of the device are hard to predict.

8.6.2.2 Statistical Approach for ESD and Reliability Parameters Verification

Integrated electronic components are required to pass standard ESD tests. Long-term reliability parameters (for example, the middle time before failure) of the components are verified by statistical methods on a limited set of components. Likewise, safe operating area (SOA) parameters are confirmed only on a limited amount of samples.

Since most of the above testing methodologies involve subjecting the device to extreme electrical overstress events, it is advised that the components subjected to the tests are no longer used in the field, even if they pass the tests. The intuitive justification for this is the possible impact of the test on long-term reliability parameters. For example, if the component has been tested for pulsed SOA or zapped with an ESD tester, then it incurs significant damage. In the case of power devices, the damage might often be latent and can be detected by measurements of some secondary parameters that produce different initial *I*–*V* characteristics. Thus, the stressed component may pass the parametric test within the datasheet specification limits while harboring latent damage.

Thus, the verification of ESD robustness and reliability parameters is provided by statistical methods applied to similar devices, while only nominal datasheet electrical parameters are verified for the devices shipped to customers.

To the best of our knowledge, no technology exists that would enable a repeated diagnostic of such electrical damages, since the impact of a sequence of diagnostic tests on the operational lifespan and long-term reliability of the device is not yet well understood. Only a statistical sampling of a group of devices can currently be done using existing technologies.

It is logical to assume that if during application or assembly a discrete device will experience an ESD event above a certain critical level, which will form induced latent damage in the device structure, then the device's long-term operation and individual time to failure may be impacted. In particular, the individual time to failure will be significantly different from the value measured by statistical methods on non-stressed devices.

Formation of the latent defects in power components is discussed below using numerical simulation analysis and results of ESD stress experiments.

8.6.3 Preliminary Numerical Analysis for Devices with Defects and the Two-Transistor Model

A numerical simulation model for the device has been used to demonstrate an effect of latent defect formation [155]. To represent the latent damage of the device, a superposition of two device structures with different width scaling factors has been used. The first device has a large width scaling factor and represents the unchanged part of the array. The second device, with a small scaling factor, had some deviation of the original parameters. The characteristics of the two-transistor mixed-mode circuit (Fig. 8.27) have been used as a physical representation of the power device array with latent damage or a local process technology defect. The last case assumes that the deviation of the device parameters is due to process technology defects.

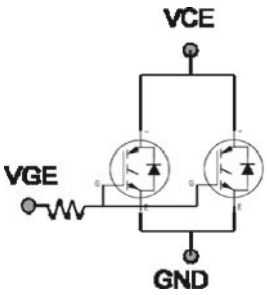


Fig. 8.27 Physical representation of the device with local defects using a two-transistor mixed-mode circuit

The physical model (Fig. 8.27) enables a more comprehensive understanding of the damage caused by ESD and is further used for comparison with experimental data in the following section.

For simplification of the numerical analysis, an isothermal case is considered, thereby ignoring the local temperature rise during ESD stress. The upper and lower part of the 2D device cross-section is shown in Fig. 8.28a, b. It displays architecture typical for a vertical gate (trench gate) IGBT cell. The upper region (Fig. 8.28a) of the device has a shallow source and body n^+ and p^+ diffusion regions, along with the second p-body diffusion that forms the well of the vertical N-channel MOS device. The poly-silicon gate is separated from the silicon by a thin gate oxide. The right side of the upper cross-section (Fig. 8.28a) represents the active device, while the left part represents the lateral termination of the array. Figure 8.28b shows the lower part of the device, which forms the heavily doped p^{++} collector region separated from Nepi region by the n-buffer.

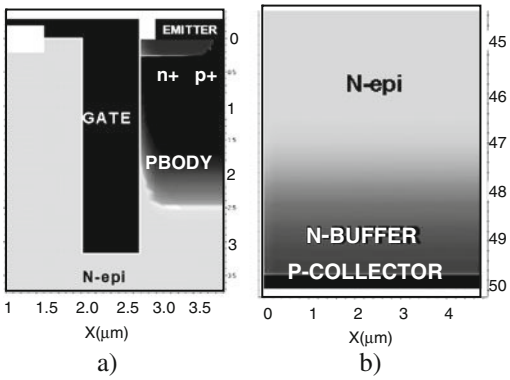


Fig. 8.28 Simplified simulation cross-section for the (**a**) upper and (**b**) lower part of the IGBT device used for numerical analysis of the experimental test data

The principle of device operation at switching is based upon the formation of an accumulation channel in the MOS structure at positive gate bias, which provides electron current through the epi region followed by hole injection from the p-collector.

In a more physically correct way, the p-collector should be called the emitter. Historically, the current terminology came in place because IGBT devices have been used as replacements for vertical double-diffusion MOSFET devices, where the n^+-drain region that connects the epi-layer at the bottom represents the collector of the parasitic NPN structure.

The controlled conductivity modulation of the Nepi region provides a rather high-current level above the typical DMOS levels, making the IGBT device an important component of power electronics.

The switching operation of the numerical IGBT model (Fig. 8.29) delivers the expected switching behavior of IGBT devices.

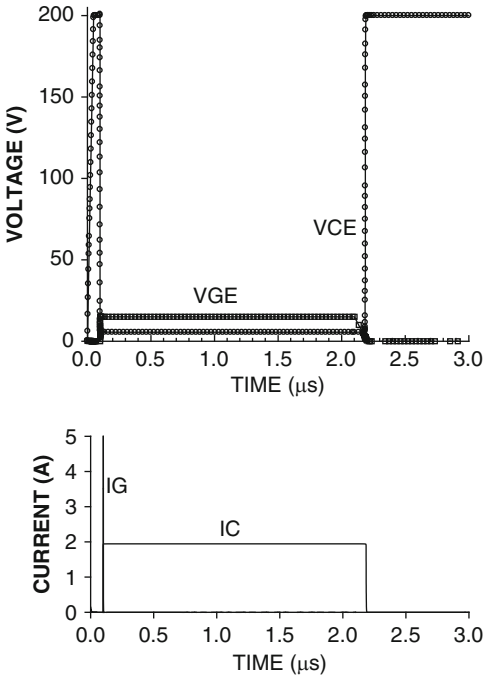


Fig. 8.29 Switching characteristics of the standard devices

Output characteristics for devices with different epi doping levels at a 5 V collector voltage are shown in Fig. 8.30c. They demonstrate a threshold voltage of ~0.9 V and a saturation current i_C of ~5e^{-5} A.

The transconductance characteristics of the device for different P_{BODY} characteristic lengths are shown in Fig. 8.30b and demonstrate the effect of the change of the threshold voltage upon this parameter.

The two-transistor model (Fig. 8.27) of the device with a latent defect can be represented by a superposition of a cross-section with original parameter $X_{PBODY} = 0.1\ \mu m$ and a cross-section with parameter $X_{PBODY} = 0.075\ \mu m$ (Fig. 8.30b). Such defect can be artificially formed as a result of ESD event. To compose a physical equivalent of the defective device, the two-transistor model incorporates components with different scaling factors.

The change of the transconductance characteristic is presented in Fig. 8.32. Depending on the width scaling factor for both cross-sections, a corresponding peculiarity will be observed on the two-transistor model transconductance characteristics.

If such a defect is formed, the local region with $X_{PBODY} = 0.075\ \mu m$ will remain in the on-state when the major part of the array is in the off state, thus providing conditions for an accelerated degradation process. The mixed-mode simulation case for a defective device can be represented by a cross-section with a width scaling

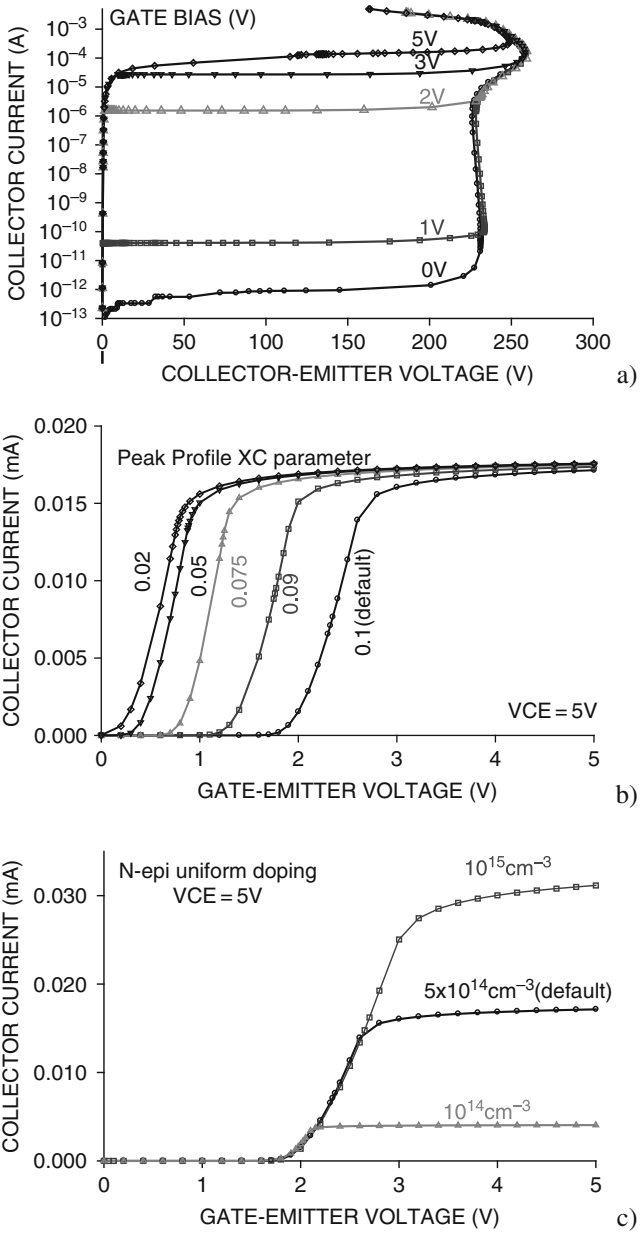


Fig. 8.30 Output I_C–V_{CE} characteristics at different gate bias of the initial device (**a**) and change of the original transconductance I_C–V_{GE} characteristics upon different characteristic PBODY profile lengths (**b**) and N-epi doping levels (**c**)

factor w_i, and an additional cross-section with the parameter deviation for defective devices w_d.

Similarly, change in other parameters of the structure results in change of the simulated characteristics. For example, change in the N-epi region (Fig. 8.30c) provides no change in the threshold voltage, while the saturation current is significantly changed. Deviation due to either process technology defects or ESD damage of the local doping levels only from 5×10^{14} cm^{-3} to 1×10^{15} cm^{-3} results in corresponding change of the current.

Another critical impact of local defects on device reliability is the change of the safe operating area (SOA). Output I_C–V_{CE} characteristics of devices with peak doping level change in the n$^+$-buffer region are presented in Fig. 8.31. When the original peak doping level of 1×10^{19} cm^{-3} in the buffer region is changed to 1×10^{18} cm^{-3}, a significant reduction of critical voltages in the whole range of gate bias is observed (Fig. 8.31), compared to the characteristics of the original device (Fig. 8.30a). In the case of the two-transistor model for defective devices, the small width scaling factor of the structure will determine the limitations of the whole device SOA.

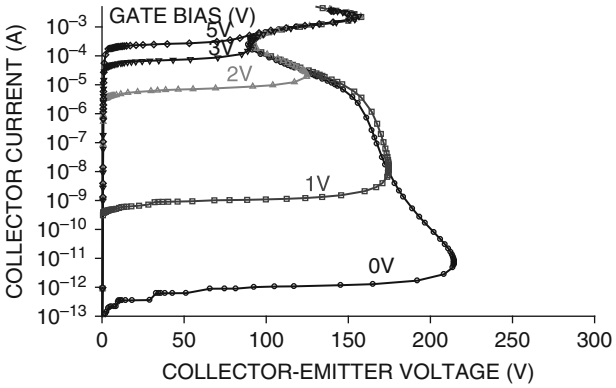


Fig. 8.31 Output I_C–V_{CE} characteristics of devices with N-epi doping deviation

The effects of local faults, i.e., localized N-epi doping deviations, are obtained through a mixed-mode simulation of the two-transistor model. An example of the change in output and transconductance characteristics for an original device with a P_{BODY} defect is presented in Fig. 8.32.

8.6.4 Experimental Evaluation of Discrete Components Robustness

The device chosen for experimental stress testing is the IRG4BC30KD. This device has a typical structure of a vertical gate (trench gate) IGBT cell (Fig. 8.33), with a

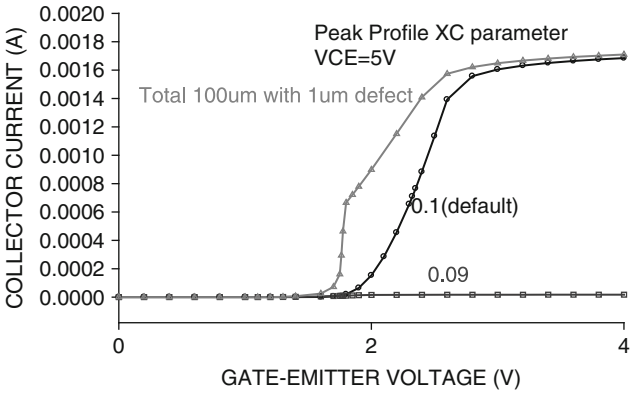


Fig. 8.32 Two-transistor model analysis: change in transconductance of 100 μm faulty device with 1 μm localized P_{BODY} doping deviation

package-integrated flyback diode. The Figure shows two separate dies, one of which contains the IGBT, while the other forms the protective diode.

8.6.4.1 TLP Stress

At first TLP measurement methodology has been applied to determine the gate ESD robustness. The gate–emitter TLP characteristics are presented for positive and negative stress (Fig. 8.34).

During TLP testing, the gate oxide is overstressed, but no change in the leakage current is observed until the critical current is ~2–3 A. No gradual changes in the device structure are observed until this level of current is reached. At the critical current level, the device fails irreversibly. The catastrophic nature of both positive and negative TLP gate overstress can be explained by the discharge of the whole charge accumulated in the gate–emitter MOS capacitor through the local damage area formed in the dielectric.

The device holds approximately three times the amount of voltage specified by the maximum limit rating, with a pulsed current that at least satisfies package-level ESD requirements.

The experimental results of catastrophic failure are repeatable for a 50-Ω TLP load.

At the same time, the source–drain TLP characteristics in discrete power components are rather robust (Fig. 8.34). In the case of the HEXFET, the device has a rather strong body diode that provides for an avalanche current above the channel saturation current in pulse conditions (Fig. 8.34). Similar avalanche energy robustness in the IGBT component is achieved by means of an integrated high-voltage Schottky diode (Fig. 8.35).

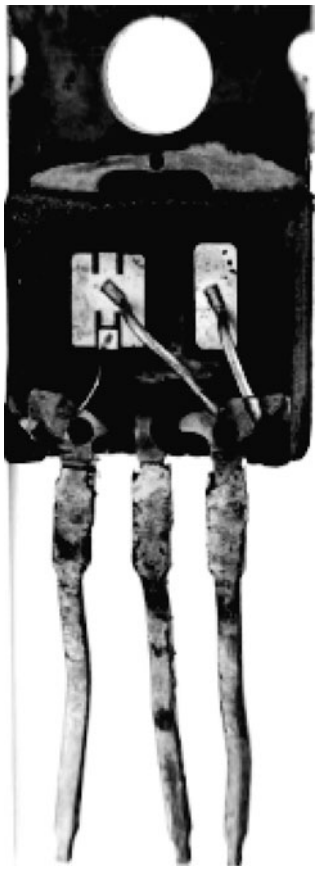


Fig. 8.33 De-capped view of the IGBT die

8.6.4.2 ISO System-Level Pulse Test

ISO gun is used in order to overcome TLP system and package-level limitations during the collector–emitter ESD stress test of the IGBT device. Both modes of operation of the gun – direct electrical contact discharge and air-gap discharge – have been used.

For the initial pool of samples, the leakage level observed after ESD zapping was consistently lower by several orders (0.01 μA vs. 250 μA) than the maximum limit of the datasheet parameters. The breakdown voltage varied around two different levels, 650 and 665 V at room temperature, above the datasheet maximum rating limit of 600 V. No failures were observed below a certain ESD pulse threshold level, even when up to 100 pulses were applied.

ESD pulse was applied to the collector–emitter junctions. The reverse current paths in IGBTs are protected by the package-level integrated diode and hence can withstand a high level of electrical voltage. The diode can withstand stress up to

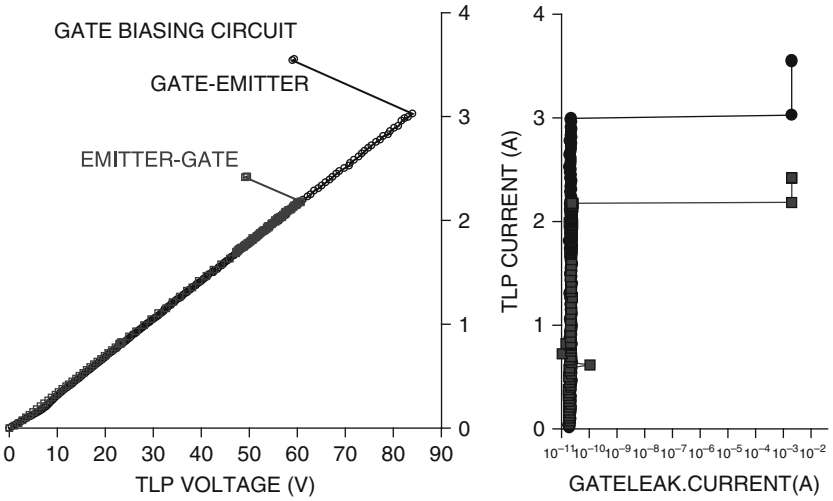


Fig. 8.34 Positive-pulsed (TLP) gate–emitter characteristics of IGBT devices subjected to gate ESD overstress

the 30 kV limit of the ISO gun. Most of the exhibited collector–emitter leakage is at hundreds of microamps to several milliamps, while most of the devices still had breakdown voltages above 600 V; several devices also showed reduction in breakdown voltages.

Although the device exceeded the datasheet leakage current limit ($\sim$250 μA) by a factor of 3, the samples remained fully functional with no additional gate leakage current. Primarily, two typical modes of fault generation were observed for ESD pulses above a certain threshold level. These involved localized defect formations that caused deviation in device parameters but no catastrophic failures. These defect modes are described in the following section.

In the case of contact zap, the gate–emitter (G–E) ESD pulse combination damaged the device irreversibly and catastrophically. The gate collector zap combination (G–C) provided parametric failure.

8.6.4.3 Collector–Emitter ISO Test

Similarly, collector–emitter ISO ESD stress above some critical limit results in either deviation of the previously measured electrical characteristics within datasheet limits, or parametric failure over the datasheet limits or catastrophic irreversible failure up to complete damage of device functionality.

Comparison of the electrical data before and after the C–E zap is presented in Fig. 8.36. As expected, the majority of the samples have leakage-breakdown characteristics that correspond to the formation of a localized region with a current level of $\sim$1 mA. This localized defect can form in the vertical emitter–collector junction

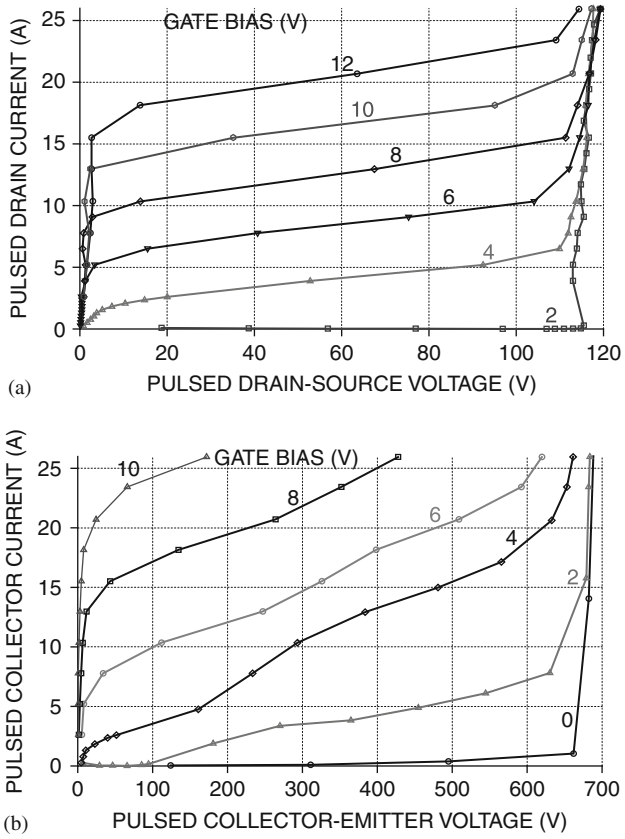


Fig. 8.35 Output TLP characteristics for discrete HEXFET with avalanche energy 90 mJ (**a**) and 16 A 600 V discrete IGBT with integrated Schottky diode (**b**)

between either the P_{BODY}-N-epi regions, or it can correspond to the damage of the n-buffer region.

Two typical modes of non-catastrophic parametric failure were observed for pulses above a certain ESD threshold level.

Characteristics for the first mode of parametric deviation within datasheet limits are shown in Fig. 8.36a, b. In the first mode, change in the device output collector–emitter characteristics occurs only after a critical voltage of ~300 V (Fig. 8.36a) and, at the same time, a part of the transconductance characteristics represents a local change in the threshold voltage (Fig. 8.36b).

Combining this experimental data with the numerical simulation analysis of the two-transistor model, this first type of damage is expected in the CMOS body region, which changes certain voltage-dependent blocking capabilities of the CMOS part of the device.

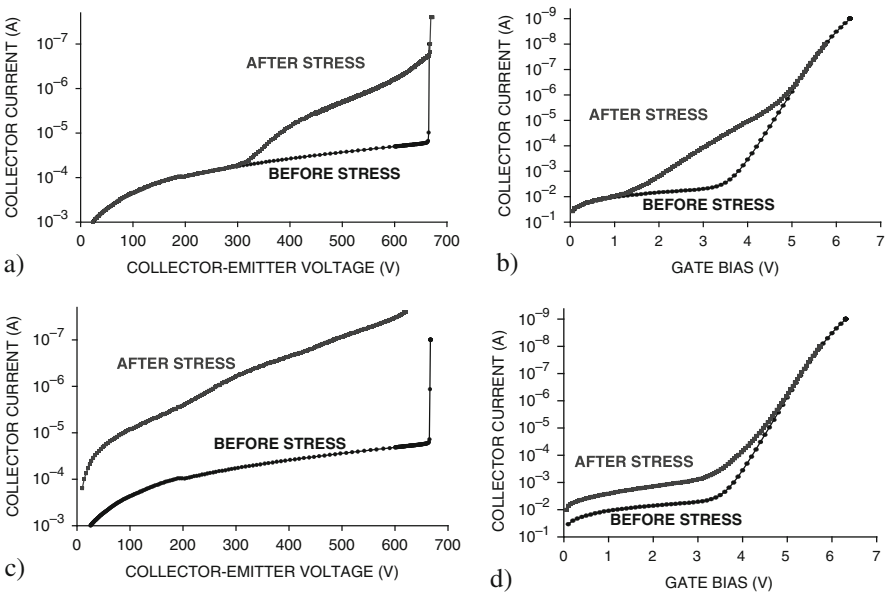


Fig. 8.36 Comparison of the collector–emitter output characteristics for first mode of non-catastrophic failure in the first (**a**) and second (**c**) mode of parametric failure and the corresponding comparison of the transconductance characteristics (**b**) and (**d**) for the same device

A different change of output characteristics is observed (Fig. 8.36c, d) for the second mode of parametric deviation within datasheet limits, where no change in the threshold voltage (Fig. 8.36d) is accompanied by a change in the leakage current in the whole voltage range. In this case, a local leakage can be assumed through the peripheral termination region of the device.

Parametric failures above datasheet limits also result in change of the SOA characteristics when the critical voltage of the whole device is below 600 V.

8.6.4.4 Gate–Collector ISO Tests

For gate–collector stress, the comparison of electrical test (ET) data before and after ESD stress leads to the conclusion that the failure signature is completely different from C–E stress. The ET data have been collected at a ~1 mA peak current limitation, based upon curve trace data. The output characteristics are shown in Fig. 8.37. Most of the samples have elevated leakage current due to the current path through the gate, while the breakdown voltage remains unchanged. This indicates that the MOS structure, rather than the collector junction, has incurred damage.

The level of change in the collector current, as well as in the transconductance, indicates more severe damage as compared with the results obtained from ESD zaps applied to the CE junction.

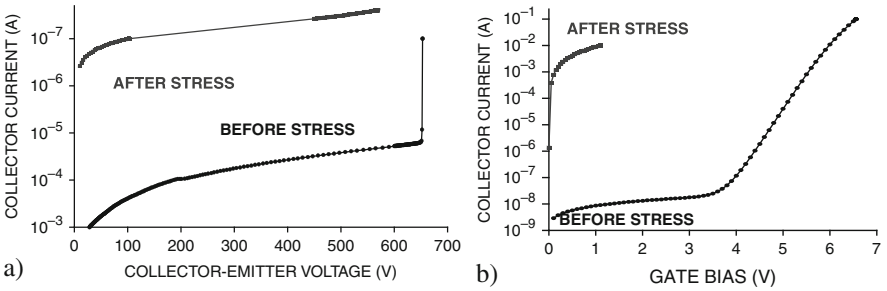


Fig. 8.37 Change of the (**a**) output characteristics and (**b**) transconductance as a result of gate–collector air-gap stress

Thus, for the case of discrete components in a certain range of ESD pulse magnitude, latent damage is observed prior to catastrophic failure both as a deviation of the ET parameters within the datasheet limits and as a change of the electrical parameters above the datasheet limits, under the preserved functionality of the stressed components. When the history of the device is unknown, previously experienced ESD stress events may result in latent damage that acts as a local defect in the device. In spite of the full functionality of the device, these damages may result in the change of SOA or the time to failure of the device in application regimes, due to the possible acceleration of the degradation processes in the region of the latent defect.

8.7 Summary

Robust ESD systems have to be insensitive to ESD discharge during operation. This robustness needs to be realized by the combination of EMC design methods both on-board and on-chip.

In addition, packaging and handling requires protection against ESD damage. This damage should be distinguished between events within EPA, where on-chip protection is sufficient in most cases, and outside EPA, where the IC is already mounted and embedded in a component.

In effect, positive IC (package) HBM or CDM stress results can be completely inadequate as criteria for designing a robust ESD system.

A careful characterization of the on-chip and on-board protection is a necessary part of system-level ESD robustness. To achieve an optimum IC design for robust ESD systems, a co-design should be performed including both chip-level and board-level protection. Finally, sufficient latch-up robustness even beyond JEDEC requirements should be provided at the exposed pins, especially for life-critical systems. At the same time, both overshoot and undershoot at supply lines should be minimized during ESD.

For a few exposed pins, a significantly increased ESD robustness is required and can often be provided on the chip using a specially designed ESD cells. On-chip system-level ESD solutions are based upon the same principles as package-level ESD protection devices, with specific attention to width scaling and the measures providing a holding voltage above the power supply level.

DECIMM^{TM} Simulation Examples for Chapter 8

To download a trial version of the numerical simulation software and request an electronic license key please visit http://www.analogesd.com
To download libraries with simulation examples for this chapter please visit http://www.analogesd.com/Chapter8.html
List of examples is subject to change.

Example 8.1 System HMM Pulse Simulation

Library Name: Examples8_System_Level_and_Discrete_Components
Project Name: E8.1_HMM

The example provides an equivalent circuit for human–machine model (HMM) ESD stress. This stress is considered as an equivalent of the system-level stress. Mixed-mode examples with NLDMOS-SCR (Fig. E8.1) demonstrate the specific features of the HMM pulse waveform and can also be used to explore system-level pulsed operation of different snapback clamps from the examples above.

Example 8.2 HMM Simulation with PCB Components and TVS

Library Name: Examples8_System_Level_and_Discrete_Components
Project Name: E8.2_HMM_with_PCB_and_TVSuppressor

The example demonstrates interaction of the parasitic printed circuit board (PCB) components L2–L4 with the snapback ESD device, resulting in voltage suppression at the input IC pin. The voltage waveforms observed at different circuit nodes are presented in Fig. E8.2, demonstrating significant reduction of the voltage amplitude at the internal system pin.

Example 8.3 Power Switch

Library Name: Examples8_System_Level_and_Discrete_Components
Project Name: E8.3a_DMOS_Switch; E8.3b_IGBT_Switch

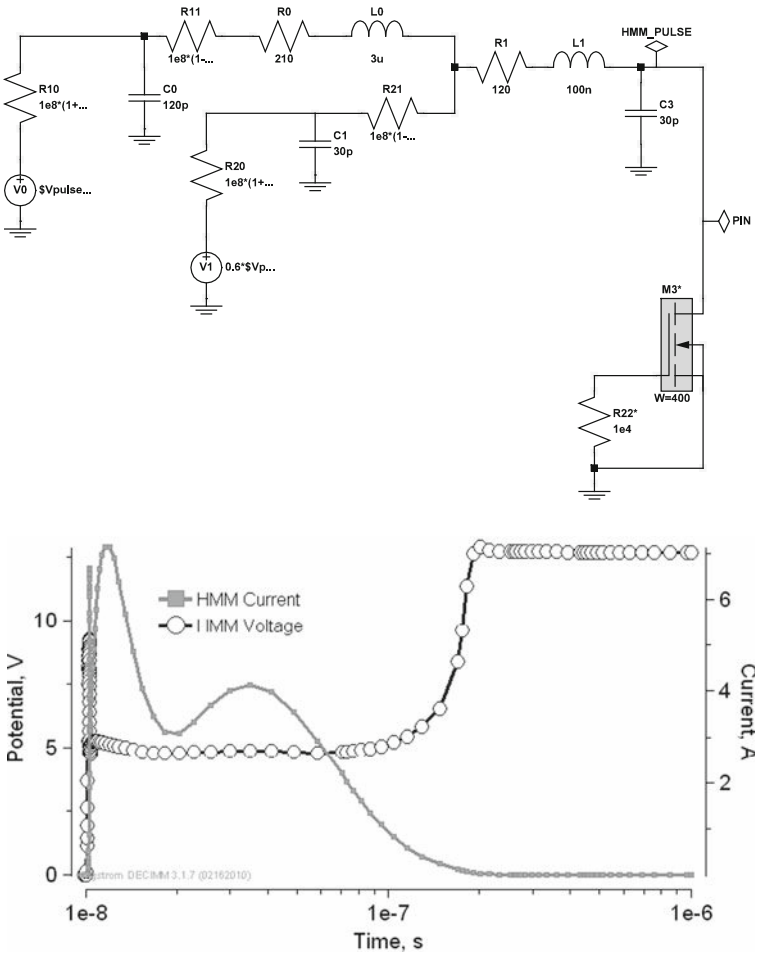


Fig. E8.1 Mixed-mode simulation circuits for an HMM pulse applied to NLDMOS-SCR with current and voltage waveforms observed at 16 kV pulse amplitude

These examples provide a basic mixed-mode simulation circuit for VDMOS and IGBT switches operating under voltage VIN of up to 600 V. The voltage and current waveforms observed during turn-on are shown in Fig. E8.3.

References

1. Amerasekera A, Duvury C (1995) ESD in Silicon Integrated Circuits. Wiley, West Sussex.
2. Dabral S, Maloney TJ (1998) Basic ESD and I/O Design. Wiley, West Sussex.
3. Estmark K, Gossner H, Stadler W (2003) Advanced Simulation Methods for ESD Protection. Elsevier, Amsterdam.
4. Voldman SH (2006) ESD RF Technology and Circuits. Wiley, Chichester.
5. Voldman SH (2004) ESD Physics and Devices. Wiley, Chichester.
6. Voldman SH (2005) ESD Circuits and Devices. Wiley, Chichester.
7. Vashchenko VA, Hopper P (2005) Bipolar SCR ESD devices. Microelectron Reliab 45: 457–471.
8. Proceedings of EOSESD Symposiums (1997–2007). ESD Association, www.esda.org
9. Vashchenko VA, Sinkevitch VF (2008) Physical Limitations of Semiconductor Devices. Springer, Berlin.
10. Stadler W (2007) HBM MIL883-STD, Method 3015.7; State-of-the-Art in ESD Standards. 1st International ESD Workshop, South Lake, Tahoe, 127–147.
11. ESD Association (1994) HBM-ESD Sensitivity Testing: Human Body Model (HBM)-Component Level, ESD Association Standard, S5.1.
12. ESD Association (1994) MM-ESD Sensitivity Testing: Machine Model (MM)-Component Level, ESD Association Standard, S5.2.
13. IEC 61000-4-2 (2008-12) Electromagnetic Compatibility (EMC) – Part 4-2: Testing and Measurement Techniques – Electrostatic Discharge Immunity Test.
14. Sze S (1981) Physics of Semiconductor Devices. Wiley, New York, NY.
15. Vashchenko VA, Sinkevitch VF (1996) Current instability and burnout of HEMT structures. Solid-State Electron 39:851–856.
16. Vashchenko VA, Kozlov NA, Martynov YB, Sinkevitch VF, Tager AS (1996) Negative differential conductivity and isothermal drain breakdown of the GaAs MESFET. IEEE Trans Electron Dev 43: 513–518.
17. Vashchenko VA, Martynov YB, Sinkevitch VF (1996) Physical limitation on drain voltage of power PM HEMT. Microelectron Reliab 37:1137–1141.
18. Vashchenko VA, Martynov YB, Sinkevitch VF, Tager AS (1996) Current instability in GaAs n^+-i-n^+ structures as a limitation of the maximum drain voltage of power MESFET's. Solid-State Electron 39:1027–1031.
19. Vashchenko VA, Martynov YB, Sinkevitch VF, Tager AS (1996) Electrical current instability at gate breakdown in GaAs MESFET. IEEE Trans Electron Devices 43(12):2080–2084.
20. Vashchenko VA, Martynov YB, Sinkevitch VF (1997) Electrical instability and filamentation in ggMOS protection structures. Solid-State Electron 41(11):1761–1767.
21. Tazzoli A, Danesin F, Zanoni E, Meneghesso G (2007) ESD robustness of AlGaN/GaN HEMT devices, 29th Electrical Overstress/Electrostatic Discharge Symposium (EOS/ESD), EOS/ESD 16–21 September 2007. 4A.4-1–4A.4-9. Anaheim, California, USA.
22. Sleiman A, Di Carlo A, Lugli P, Meneghesso G, Zanoni E, Thobel JL (2003) Channel thickness dependence of breakdown dynamic in InP-based lattice-matched HEMTs. IEEE Trans Electron Devices 50(10):2009–2014.

V.A. Vashchenko, A. Shibkov, *ESD Design for Analog Circuits*,
DOI 10.1007/978-1-4419-6565-3, © Springer Science+Business Media, LLC 2010

23. Bonch-Bruevitch VL, Zviagin NP, Mironov AG (1975) Domain Electrical Instability in Semiconductors. Nauka, Moskow.
24. Nikolis R, Prigogine I (1979) Self-organization in nonequilibrium systems. Mir, Moscow.
25. Haken H (1980) Sinergetics. Mir, Moscow.
26. Kerner BS, Osipov VV (1994) Autosolitons. A new approach to Problems of Self-Organization and Turbulence. Kluwer Academic Publishers, Dordrecht, Boston, London.
27. Scholl E (1987) Nonequilibrium Phase Transition in Semiconductors Self-organization Induced by Generation and Recombination Process. Springer, Heidelberg.
28. Moens P, Bychinkin S, Reynders K, et al. (2004) Effect of hot spot hopping and drain ballasting integrated vertical DMOS devices under TLP stress. Proc. Int. Rel. Phys. Symp., 393–398. Phoenix, AZ, USA.
29. Gossner H, Muller-Lynch T, Eastmark K, Stecher M (1999) Wide range control of sustaining voltage of ESD protection elements realized in smart power technology. Proc. EOS/ESD Symp., 19–27. Phoenix, AZ, USA.
30. Keppens B, Mergens M, Armer J, et al. (2003) Active-area-segmentation (AAS) technique for compact ESD robust, fully silicided NMOS design. Proc. EOS/ESD Symp., 250–258.
31. DECIMM^{TM} Angstrom Design Automation (2010) Release 2.0, www.analogesd.com
32. Hanwa (2009) HBM, MM Waveform Capture System. Hanwa Electronic IND. Co., Ltd., Japan.
33. Scholz M, et al. (2007) Calibrated wafer-level HBM measurements for quasi-static and transient device analysis. Proceedings of EOS/ESD Symposium, 89–94.
34. Grutchfield HB, Moutoux TJ (1966) Current mode second breakdown in epitaxial planar transistors. IEEE Trans Electron Devices 13:743–747.
35. Oxner ES (1985) Power Field Effect Transistor and Its Application. Prentice-Hall, Englewood Cliffs, NJ.
36. Vashchenko VA, Kuznetsov V, Hopper P (2007) Implementation of dual-direction SCR devices in analog CMOS process. Proc. EOSESD Symp., 75–79.
37. Concannon A, Vashchenko VA, ter Beek M, Hopper P (2003) A device level negative feedback in the emitter line of SCR-structures as a method to realize latch-up free ESD protection. 41st Annual 2003 IEEE International Reliability Physics Symposium Proceedings, 30 March–4 April 2003, 105–111 .
38. Russ C, Mergens M, Verhaege K, et al. (2001) GGSCRs: GGNMOS triggered silicon controlled rectifiers for ESD protection in deep sub-micron CMOS processes. Proc. EOS/ESD Symp., pp. 22–31.
39. Wolf S (1990) Silicon Processing for VLSI Era, Vol. 2, Process Integration. Lattice Press, Sunset Beach, CA.
40. LaFonteese D, Vashchenko VA, Korablev KG (2009) Breakdown voltage walkout effect in ESD protection devices. IRPS.
41. Wu J, Rosenbaum E (2004) Gate oxide reliability under ESD-like pulse stress. IEEE Trans Educ 51(7):1192–1195.
42. Vashchenko VA, Kindt W, Hopper P, ter Beek M (2004) Implementation of 60 V tolerant dual direction ESD protection in 5V BiCMOS process for automotive application. Proc. EOS/ESD Symposium.
43. Wunsch D, Bell R (1968) Determination of threshold failure levels of semiconductor diodes and transistors due to pulse voltages. IEEE Trans Nucl Sci NS-15:244–259.
44. Dwyer V, Franklin A, Campbell D (1991) Thermal failure in semiconductor devices. Solid-State Electron 33:553–560.
45. Tasca D (1970) Pulsed power failure modes in semiconductors. IEEE Trans Nucl Sci NS-17:364–372.
46. Charitat G (2001) Voltage handling capability and termination techniques of silicon power semiconductor devices. IEEE BCTM 10.5, 175–183.
47. Vashchenko VA, ter Beek M (2005) ESD protection window targeting using LDMOS-SCR devices with PWELL-NWELL super-junction. Proc. IRPS, April 2005, 612–613.

48. Chung Y, Xu H, Ida R, Baird R (2006) Snapback breakdown dynamics and ESD susceptibility of LDMOS. Proc. Int. Rel. Phys. Symp., 352–355; Steinhoff R, Huang J-B, Hower P, Brodsky J (2003) Current filament movement and silicon melting in an ESD-robust DENMOS transistor. Proceed. EOS/ESD Symp., 220–268.
49. Vashchenko VA, Lafonteese D (2009) System level and hot plug-in protection of high voltage transient pins. Proc. EOS/ESD Symposium.
50. Vashchenko VA, Concannon A, ter Beek A, Hopper P (2002) Technology CAD evaluation of BiCMOS protection structures operation including spatial thermal runaway. Proc. EOSESD Symposium, 101–110.
51. Chatterjee A, Polgreen T (1991) A low voltage triggering SCR for on-chip ESD protection at output and input pads. IEEE Electron Device Lett 12(1):21–22.
52. Chen JZ, Amerasekera A, Vrotos T (1995) Bipolar SCR ESD protection for high speed submicron bipolar/BiCMOS circuit frequency integrated circuits. IEDM, 337–340.
53. Vashchenko VA, LaFonteese D (2008) Lateral PNP BJT ESD protection devices. BCTM, 53–56.
54. Li J, Gauthier R, Chatty K, Kontos D, et al. (2005) PMOSFET-based ESD protection in 65 nm bulk CMOS technology for improved external latch-up robustness. Proc. EOS/ESD Symposium.
55. Gendron A, Renaud P, Besse P, Salamero C, Bafleur M, Nolhier N (2006) Area-efficient reduced and no-snapback PNP-based ESD protection in advanced smart power technology. Proc. EOS/ESD Symposium, September 2006, 69–76.
56. Renaud P, Gendron A, Bafleur M, Nolhier N (2007) High robustness PNP-based structure for the ESD protection of high voltage I/Os in an advanced smart power technology. Proceed. BCTM, 226–229.
57. Vashchenko VA, LaFonteese D (2009) System level and hot plug-in protection of high voltage transient pins. EOS/ESD Symposium.
58. Voldman S, Giessier S, Nakoe J, et al. (1998) Semiconductor process and structural optimization of shallow trench isolation-defined and polysilicon-bound source/drain diodes for ESD networks. Proc. EOS/ESD Symp., 151–160.
59. Concannon A (2005) National Semiconductor, Internal Engineering Analysis Report.
60. Merrill R, Issaq E (1993) ESD design methodology. Proc. ESD/EOS Symposium, 233–237.
61. Stockinger M (2003) Boosted and distributed rail clamp networks for ESD protection in advanced CMOS technologies. EOS/ESD Symposium Proceedings.
62. Anderson W, Krakauer D (1998) Cross referenced ESD protection for power supplies. Proc. EOS/ESD Symp.
63. Croft G (1996) Transient supply clamp with variable RC time constant. Proc. EOS/ESD Symp.
64. Dabral S, Aslett R, Maloney T (1994) Core clamps for low voltage technologies. Proc. EOS/ESD Symp.
65. Juliano P, Anderson W (2003) ESD protection design challenges for a high pin-count alpha microprocessor in a 0.13um CMOS SOI technology. Proc. EOS/ESD Symp.
66. Li J, Gauthier R, Rosenbaum E (2004) A compact, timed shut-off, MOSFET-based power clamp foe on chip ESD protection. Proc. EOS/ESD Symp.
67. Poon SS, Maloney T (2002) New consideration for MOSFET power clamps. Proc. EOS/ESD Symp., 1–5.
68. Stockinger M, Miller J (2004) Advanced ESD rail clamp network design for high-voltage CMOS applications. Proc. EOS/ESD Symp.
69. Worley ER, Gupta R, Jones B, et al. (1995) Sub-micron chip ESD protection schemes which avoid avalanching junctions. Proc. EOS/ESD Symp., 13–20.
70. Maloney T, Kan W (1999) Stacked PMOS clamps for high voltage power supply protection. Proc. EOS/ESD Symp., 70–77.
71. Smedes T, Heringa A, van Zwol J , de Jong P (2002) The impact of substrate resistivity on ESD protection devices. Proc. EOS/ESD Symp., 354.

72. Vashchenko VA, Concannon A, ter Beek M, Hopper P (2003) Increasing the ESD protection capability of over-voltage NMOS structures by comb-ballasting region design. 41st Annual 2003 IEEE International Reliability Physics Symposium Proceedings, 30 March–4 April 2003, 261–268.
73. Vashchenko VA, Concannon A, ter Beek M, Hopper P (2003) ESD-implant effect on protection capability of NMOS structures. 33rd Conference on European Solid-State Device Research, 2003. ESSDERC'03. 16–18 September 2003, 565–568.
74. Quittard Q, Mrcarica Z, Blanc F, Notermans G, Smedes T, van Zvol H (2006) ESD protection for high voltage CMOS technologies. Proc. EOS/ESD Symp., 77–86.
75. Vassilev V, Lorenzini M, Jansen P, Vashchenko VA, et al. (2003) Snapback circuit model for cascoded NMOS ESD over-voltage protection structures. 33rd Conference on European Solid-State Device Research, 2003. ESSDERC'03, 16–18 September 2003, 561–564.
76. Smedes T, Li Y (2003) ESD phenomena in interconnect structures. Proc. EOS/ESD Symp.
77. Boselli G, Duvvury C, Reddy V (2003) Efficient pnp characteristics of pMOS transistors in sub-0.13um ESD protection circuits. Proc. EOS/ESD Symp., 257–266.
78. Duvvury C, Ramaswamy S, Amerasekera A, Cline RA, Andresen BH, Gupta V (2000) Substrate pump NMOS for ESD protection applications. Electrical Overstress/Electrostatic Discharge Symposium Proceedings, 26–28 September 2000, 7–17.
79. Vashchenko VA, Concannon A, ter Beek M, Hopper P (2002) Emitter injection control in LVTSCR for latch-up free ESD protection. Proc. MIEL 2:741.
80. Meneghesso G, Tazzoli A, Marino FA, Cordoni M, Colombo P. Development of a new high holding voltage SCR-based ESD protection structure. IEEE International Reliability Physics Symposium, IRPS 2008, April 27–May 1 2008, 3–8.
81. Vashchenko VA, Concannon A, ter Beek M, Hopper P (2004) High holding voltage cascoded LVTSCR structures for 5.5-V tolerant ESD protection clamps. IEEE Trans Device Material Reliab 4(2):273–280.
82. Synopsys (2002) "TSUPREM-4" Two-Dimensional Process Simulation Program, Version 2002.4.0.
83. Synopsys (2002) MEDICI Two-Dimensional Device Simulation Program, Version 2002.4.0.
84. Chatty K, Peachey N, Quek C, et al. (2010) SCR-based power clamp for low leakage applications. Proceedings of the International ESD Workshop, Poster 2.7.
85. Vashchenko VA, Farrenkopf D, Hopper P (2007) Active control of the triggering characteristics of NPN BJT, BSCR and NLDMOS-SCR devices. 19th International Symposium on Power Semiconductor Devices and IC's, ISPSD'07, 27–31 May 2007, 41–44.
86. Avery L (1983) Using SCR's as transient protection structures in integrated circuits. Proceedings of ESD/EOS Symposium, 27–29.
87. Mergens M, Russ C, Verhaege K, et al. (2003) Diode-triggered SCR (DTSCR) for RF-ESD protection of BiCMOS SiGe HBTs and CMOS ultra-thin gate oxides. Intl. Electron Devices Meeting Technical Digest, 515–518.
88. Di Sarro J, Vashchenko VA, Rosenbaum E, Hopper P (2008) A dual-base triggered SCR with very low leakage current and adjustable trigger voltage. 30th Electrical Overstress/Electrostatic Discharge Symposium, EOS/ESD, 7–11 September 2008, 242–248.
89. Olson N, Vashchenko V, Rosenbaum E, Hopper P (2008) Small footprint trigger voltage control circuit for mixed-voltage applications. 30th Electrical Overstress/Electrostatic Discharge Symposium, EOS/ESD, 7–11 September 2008, 196–203.
90. Vashchenko VA, Olson N, Farrenkopf D, Kuznetsov V, Hopper P, Rosenbaum E (2007) Mixed device-circuit solution for ESD protection of high-voltage fast pins. 45th Annual IEEE International Reliability Physics Symposium, Proceedings, 15–19 April 2007, 602–603.
91. Di Sarro J, Chatty K, Gauthier R, Rosenbaum E (2007) Evaluation of SCR-based ESD protection in 90nm and 65nm CMOS technologies. Proc. Intl. Reliability Physics Symp., 348–357.

92. Tan PY, Indrajit M, Li PH, Voldman S (2005) RC-triggered PNP and NPN simultaneously switched silicon controlled rectifier ESD networks for sub-0.18 μm technology. Proc. Intl. Symp. on the Physical and Failure Analysis of Integrated Circuits, 71–75.
93. Gerdemann A, Bhatia K, Rosenbaum E (2007) A Kelvin transmission line pulsing system with optimized oscilloscope ranging. Proc. EOS/ESD Symp., 80–88.
94. Chen J, Zhang X, Amerasekera A, Vrotsos T (1996) Design and layout of a high ESD performance NPN structure for submicron BiCMOS/bipolar circuits. Proc. Int. Rel. Phys. Symp., 227–232.
95. Coffing D, Ida R (1998) Analysis of a Zener-triggered bipolar ESD structure in a BiCMOS technology. Proc. BCTM, 227–232.
96. Vashchenko VA, Kuznetsov V, Hopper P (2007) ESD protection of fast transient pins in bipolar processes. Bipolar/BiCMOS Circuits and Technology Meeting, BCTM'07, IEEE, 222–225.
97. Chen JZ, Amerasekera A, Vrotos T (1995) Bipolar SCR ESD protection for high speed submicron bipolar/BiCMOS frequency integrated circuits. IEDM, 337–340.
98. Vashchenko VA, Hopper P (2006) A new principle for a self-protecting power transistor array design. Power Semiconductor Devices and IC's, ISPSD 2006. IEEE International Symposium on 4–8 June 2006, 1–4; Vashchenko VA, Hopper PJ (2006) Self-protecting arrays for open drain circuits. 44th Annual IEEE International Reliability Physics Symposium Proceedings, 26–30 March 2006, 637–638.
99. Trajkovic T, Udugampola N, Pathirana V, Mihaila A, Udrea F, Amaratunga GA, Koutny B, Ramkumar K, Geha S (2009) High frequency 700 V PowerBrane LIGBTs in 0.35μm bulk CMOS technology. 21st International Symposium on Power Semiconductor Devices & IC's, ISPSD 2009, 14–18 June 2009, 307–310.
100. Udrea F (2007) SOI-based devices and technologies for high voltage ICs. IEEE Bipolar/BiCMOS Circuits and Technology Meeting, BCTM'07, September 30, 2007–Oct. 2, 2007, 74–81.
101. Scholz M, Linten D, Thijs S, et al. (2010) Transient characterization of high voltage switching devices. Proceedings of International ESD Workshop, Poster 2.4.
102. Vashchenko VA, Hopper P (2005) BSCR ESD protection in 250 V process taking into account the turn-off effect. Proceedings of the Bipolar/BiCMOS Circuits and Technology Meeting, 9–11 October 2005, 70–73.
103. Linten D, Vashchenko VA, Scholz M, et al. (2008) Extreme voltage overshoots in HV snapback devices during HBM ESD stress. Proc. EOS/ESD Symp.
104. Eherton M, Khazhinsky M, Miller J, et al. (2009) EDA tool for checking ESD specific I/O ring integration rules. International ESD Workshop.
105. Concannon A (2007–2008) National Semiconductor, Internal Engineering Analysis Report.
106. Anderson WR (2009) I/O Design Concepts for the ESD Engineer, International ESD Workshop, South Lake, Tahoe, 2009 Seminar #1.
107. Joshi S, Rosenbaum E (2002) Compact modeling of vertical ESD protection NPN transistors for RF circuits. Proc. EOS/ESD Symp., 289–295.
108. Vassilev V, Groeseneken G, Jenei S, Maes H (2002) Modeling and extraction of RF performance parameters of CMOS electrostatic discharge protection devices. Proc. EOS/ESD Symp., 111–118.
109. Russ C, Verhaege K, Bock K, et al. (1996) A compact model for the grounded gate NMOS behavior under CDM ESD stress. Proc. EOS/ESD Symp., 302–315.
110. Mergens M, Wilkening W, Mettler S, et al. (1999) Analysis and compact modeling of lateral DMOS power devices under ES stress conditions. Proc. EOS/ESD Symp., 1–10.
111. Vassilev V, et al. (2001) Analysis arid improved compact modeling of the breakdown behavior of sub-0 25 micron ESD protection ggNMOS device. Proc. EOSESD Symp., 62–70.
112. Vassilev V, Vashchenko VA, Jansen P, et al. (2005) ESD circuit model based protection network optimisation for extended-voltage NMOS drivers. Microelectron Reliab 45: 1430–1435.

113. Aliaj B, Vashchenko V, Cui Q, Liou J, Tcherniaev A, Ershov M, LaFonteese D (2009) 2.5-Dimensional simulation for analyzing power arrays subject to ESD stresses. Electrical Overstress/Electrostatic Discharge Symposium Proceedings, 3A.2.1–3A.27.
114. Monticelli D (2004) The future of complementary bipolar. BCTM, 21–25.
115. Denison M, Murtaza S, Steinhoff R (2007) 25V ESD NPN transistor optimized by distributed emitter ballasting using emitter contact area segmentation. Proc. Int. Rel. Phys. Symp., 604–605.
116. Jansen P, Thijs S. Linten D, et al. (2005) RF ESD protection strategies – the design and performance trade-off challenges. Proceedings of the IEEE 2005 Custom Integrated Circuits Conference, 21–21 September 2005, 489–496.
117. Stadler W, Brodbeck T, Gartner R, Gossner H (2006) Cable discharges into communication interfaces. Proc. EOS/ESD Symposium, 144–151.
118. Pommerenke D (2001) Charged cable event. IEEE 802.3 Cable Discharge Ad-Hoc, February 2001.
119. Pischl N (2005) ESD transfer through Ethernet magnetics. Proc. Intl. EMC Symposium, 356–363.
120. Deatherage J, Jones D (2000) Multiple factors trigger cable discharge events in Ethernet LANs. Electron Design 48(25):111–116.
121. Ram C (2007) Cable Discharge Event (CDE) Stress Simulation Test Procedure for Integrated Circuits. Texas Instruments. TX, USA.
122. Sandler SM (2002) Switchmode Power Supply Simulation with Pspice and SPICE3. McGraw-Hill, New York, NY.
123. Erickson R, Maksimovic D (2001) Fundamentals of Power Electronics. Kluwer Academic Press, Dordrecht.
124. Mitchell D (1988) Dc-dc Switching Regulator Analysis. McGraw-Hill, New York, NY.
125. Basso CP (2008) Switch-Mode Power Supplies. McGraw-Hill, New York, NY.
126. Moss J, Ranucci P, Bikulcius S (2009) Optimizing power FET design. Innovation Solution Conference, National Semiconductor.
127. Goyal A, Whitfield J, Hong C, et al. (2008) Unique ESD failure mechanism of high voltage LDMOS transistors for very fast transients. IEEE IRPS, Phoenix, 673–674.
128. Chung Y, Xu H, Ida R, Min WG, Baird B (2005) ESD scalability of LDMOS devices for self-protected output drivers. Proceedings of IS PSD, 351–354.
129. Chung Y, et al. (2006) Proc. Int. Rel. Phys. Symp. 352; Steinhoff R, et al. (2003) Proc. EOS/ESD Symp., 220.
130. Parthasarathy V, Khemka V, Zhu R, et al. (2002) A multi trench analog+logic protection (M-TRAP) for substrate crosstalk prevention in a 0.25 μm smart power platform with 100V high-side capability. IEEE Electron Device Lett 23(4):212–214.
131. Vashchenko VA, et al. (2010) International ESD Workshop, in print.
132. Hower P (2010) Private Communication.
133. Mergens M, Mayerhofer MT, Willemen JA, Stecher M (2006) ESD protection considerations in advanced high-voltage technologies for automotive. Proc. EOSESD, 54.
134. Chen W, Ker M-D, Jou Y, et al. (2009) Source-side engineering to increase holding voltage of LDMOS in a 0.5um 16-V BCD technology to avoid latch-up failure. IEEE Proc. of 16th IPFA; High-voltage nLDMOS in Waffle-layout style with body-injected technique for ESD protection. IEEE Electron Device Lett 30(4):389–391.
135. Aliaj B, Vashchenko VA, et al. (2009) EOS/ESD Symp., 3A.2-7126; Moss J, Ranucci P, Bikulcius S (2009) Optimizing power FET design. Innovation Solution Conference, National Semiconductor.
136. Lafonteese D, Vashchenko VA, Linten D, et al. (2009) Self-protection capability of power arrays. Proceedings EOS/ESD Symposium.
137. Virtuoso Spectre Circuit Simulator User Guide, Version 7.0.1, June 2008.
138. Chapra SC (2006) Applied Numerical Methods with Matlab. McGraw-Hill, New York, NY. EOS/ESD placeholder for reviewer suggestion Chapter 3.

139. Gossner H (2009) ESD robust systems–IC design contribution, South Lake, Tahoe, Seminar #2.
140. Gossner H, Drüen S, Esmark K, et al. (2005) SoC. A real challenge for ESD protection. Proc. EOSESD.
141. Mergens M, Mayerhofer MT, Willemen JA, Stecher M (2006) ESD protection considerations in advanced high-voltage technologies for automotive. Proc. EOSESD, 54.
142. Bonfert D, Gieser H, Bock, K, Svasta P, Ionescu C (2009) Pulsed behavior of polymer protection devices. Electron Technol. ISSE 2009. 32nd International Spring Seminar on 13–17 May 2009, 1–6.
143. Scholz M, Linten D, Thijs S, Sawada M, Nakaei T, Hasebe T, Groeseneken G (2008) On-wafer human metal model. System-level ESD stress on component level. Proc. RCJ.
144. Scholz M, Linten D, Thijs S, Sawada M, Nakaei T, Hasebe T, Lafonteese D, Vashchenko VA, Vandersteen G, Groeseneken G (2009) On-wafer human metal model measurements for system-level ESD analysis. EOSESD Symposium.
145. Ashton R (2007) System level ESD testing – the waveforms. Conformity Magazine 12(10):28–34.
146. Stadler W, et al. (2004) From the ESD robustness of products to the system ESD robustness. Proceedings EOS/ESD Symposium, 67–74.
147. Peachey N (2008) ESD Open Forum – ESD stressing of devices and components. Conformity Magazine 13(6):22–23.
148. Wang K, et al. (2004) Characterization of human metal ESD reference discharge event and correlation of generator parameters to failure levels – part II: Correlation of generator parameters to failure levels. IEEE Trans EMC 46(4):505–511.
149. Scholz M, et al. (2007) Calibrated wafer-level HBM measurements for quasi-static and transient device analysis. Proceedings EOS/ESD Symposium, 89–94.
150. Wang AZ, Tsay CH (1999) A compact square-cell ESD structure for BiCMOS IC. IEEE BCTM 3.1, 46.
151. Vashchenko VA, ter Beek M, Kindt W, Hopper P (2004) ESD protection of the high voltage tolerant pins in low-voltage BiCMOS processes. Proceedings of BCTM, 277–280.
152. Vashchenko VA, Concannon A, ter Beek M, Hopper P (2003) Multi-port ESD protection using bi-directional SCR structures. Proc. BCTM, 137–140.
153. Besse P (2010) ESD/EMC in an automotive environment. Proceedings of International ESD Workshop, Seminar #2.
154. Paddan P, Best M, Ferrer E, et al. (2010) An ESD robust wireless sensor chip for medical application. Proceedings of International ESD Workshop, Keynote Presentation.
155. Wysocki1 P, et al. (2009) Effect of electrostatic discharge on electrical characteristics of discrete electronic components (technical brief). PHM Conference, Anaheim.
156. Davidson S (2006) Searching for clues: diagnosing IC failures. IEEE Design Test Comput 23(1):67–68.
157. Desineni R, Blanton RD (2005) Diagnostics of arbitrary using neighborhood function extraction. Proceedings of 23rd IEEE VLSI Test Symposium, 366–373; Khatir Z, Skipedj J, Lallemand R (1994) Main technological causes and some mechanisms of high power GTO's failures. Proceedings of Fifth International Conference on Power Electronics and Variable-Speed Drives, 319–324.
158. Li TS, Huang CL (2009) Defect spatial pattern recognition using a hybrid SOM-SVM approach in semiconductor manufacturing. Expert Syst Appl Int J 36(1):374–385.
159. Salman AA, Beebe SG, Emam M, Pelella MM, Ioannou DE (2006) A novel device for ESD protection in deep sub-micron SOI technologies. International Field Effect Diode (FED) Electron Devices Meeting, IEDM'06, 11–13 December 2006, 1–4.
160. Ille A, Stadler W, Kerber A, et al. (2006) Ultra thin gate oxide reliability in the ESD time domain. Proc. EOS/ESD Symp.

161. Khazhinsky M, Stockinger M, Miller J, Weldon JC (2004) Engineering single NMOS and PMOS output buffers for maximum failure voltage in advanced CMOS technolo-gies. Proc. EOS/ESD Symp.
162. Salman A, Gauthier R, Wu E, et al. (2002) Electrostatic discharge induced oxide breakdown characterization in a 0.1 um CMOS technology. 40th International Reliability Phys. Symp. Proceedings, 170–174.
163. Stathis J, Maria D (1998) Reliability projection for ultra-thin oxide reliability in CMOS devices and circuits. IEDM.
164. Weir B, Leung CC, Silverman P, Alam M (2005) Gate dielectric breakdown ion the time-scale of ESD events. Microelectron Reliab 45:427–436.
165. Hsueh KK, Sin-Hao K, Lee J, Rosenbaum E (2008) UVeriESD: an ESD verification tool for SoC design. IEEE Asia Pacific Conference on Circuits and Systems, 2008. APCCAS 2008, November 30, 2008–December 3, 2008, 53–56.
166. Vashchenko VA, Scholz M, Jansen Ph, et al. (2006) Turn-off characteristics of the CMOS snapback ESD protection devices – new insights and its implications. Proceed. EOS/ESD Symp., 39–45.
167. Worley E, Bakulin A (2002) Optimization of input protection diode for high speed applications. Proc. EOS/ESD Symp., 62–72.
168. Concannon A, Vashchenko VA, ter Beek M, Hopper P (2004) ESD protection of double-diffusion devices in submicron CMOS processes. Proc. ESSDERC, 261–264.

Index