Technical Report

UCAM-CL-TR-811
ISSN 1476-2986

Number 811

UNIVERSITY OF CAMBRIDGE

Computer Laboratory

Active electromagnetic attacks on secure hardware

A. Theodore Markettos

December 2011

15 JJ Thomson Avenue
Cambridge CB3 0FD
United Kingdom
phone +44 1223 763500

http://www.cl.cam.ac.uk/

This technical report is based on a dissertation submitted March 2010 by the author for the degree of Doctor of Philosophy to the University of Cambridge, Clare Hall.

Technical reports published by the University of Cambridge Computer Laboratory are freely available via the Internet:

http://www.cl.cam.ac.uk/techreports/

ISSN 1476-2986

Active electromagnetic attacks on secure hardware

A. Theodore Markettos

Summary

The field of side-channel attacks on cryptographic hardware has been extensively studied. In many cases it is easier to derive the secret key from these attacks than to break the cryptography itself. One such side-channel attack is the electromagnetic side-channel attack, giving rise to electromagnetic analysis (EMA).

EMA, when otherwise known as 'TEMPEST' or 'compromising emanations', has a long history in the military context over almost the whole of the twentieth century. The US military also mention three related attacks, believed to be: HIJACK (modulation of secret data onto conducted signals), NONSTOP (modulation of secret data onto radiated signals) and TEAPOT (intentional malicious emissions).

In this thesis I perform a fusion of TEAPOT and HIJACK/NONSTOP techniques on secure integrated circuits. An attacker is able to introduce one or more frequencies into a cryptographic system with the intention of forcing it to misbehave or to radiate secrets.

I demonstrate two approaches to this attack.

To perform the reception, I assess a variety of electromagnetic sensors to perform EMA. I choose an inductive hard drive head and a metal foil electric field sensor to measure near-field EM emissions.

The first approach, named the re-emission attack, injects frequencies into the power supply of a device to cause it to modulate up baseband signals. In this way I detect data-dependent timing from a 'secure' microcontroller. Such up-conversion enables a more compact and more distant receiving antenna.

The second approach involves injecting one or more frequencies into the power supply of a random number generator that uses jitter of ring oscillators as its random number source. I am able to force injection locking of the oscillators, greatly diminishing the entropy available.

I demonstrate this with the random number generators on two commercial devices. I cause a 2004 EMV banking smartcard to fail statistical test suites by generating a periodicity. For a secure 8-bit microcontroller that has been used in banking ATMs, I am able to reduce the random number entropy from 2^{32} to 225. This enables a 50% probability of a successful attack on cash withdrawal in 15 attempts.

Acknowledgments

This work would not have been possible without the support of a large number of people.

Firstly, a number of parts were done in collaboration with others. I'd like to thank the following for the contributions they made:

- Andrew West laid out the antennas for the Lochside Emissions Testing Block (ETB).
- Robert Mullins and Andrew West merged the Lochside ETB with the rest of the design and did top level routing and layout. They also arranged fabrication, designed the PCB, and wrote microcontroller firmware to control the Lochside network-on-chip (which I subsequently used as the basis of the ETB firmware).
- Markus Kuhn provided scripts for instrument control that I ported and expanded; also the idea and implementation of interpolation of time-domain traces due to variable trigger delays. He also provided the secure microcontroller board for the random number experiments, and detailed advice on that work.
- Sergei Skorobogatov designed and made the second AC280 hard drive head amplifier board. He also supplied XYZ stages, their controller, and prepared associated metalwork.
- Steven Murdoch provided a USB smartcard reader and a Python library for performing EMV transactions on banking cards, as well as useful advice on the detailed workings of the EMV protocol.

Figure 4.6 on page 85 has been reproduced from Caruso, Bratland, Smith, and Schneider (1998). Some Springbank and Lochside layouts and board photos have been taken from their documentation.

This thesis has been typeset using LaTeX, with graphics produced using the Ti*k*Z, *pgfplots* and *matlab2tikz* packages. Layout is modified from a stylesheet written by Carys Underdown. Much other open source software was used in this research and its presentation: particular examples are highlighted in the text.

This research was funded in part by the Engineering and Physical Sciences Research Council, the Computer Laboratory, Clare Hall and the Cambridge Philosophical Society.

I would like to thank the long list of people who provided ideas and comments when discussing my work, not least Drs Simon Hollis, Petros Oikonomakos, Philip Paul, Huiyun Li, Robert Mullins, Andrew West, Sergei Skorobogatov, Saar Drimer, Steven Murdoch, Markus Kuhn, the late David Moore, Professor Ross Anderson and other members of the Security and Computer Architecture groups.

I would like to thank Markus Kuhn and Elisabeth Oswald as my examiners.

Dr Bobbie Wells and Irene Hills at Clare Hall were tirelessly supportive. Dr George Reid at OIS was also instrumental in resolving an unexpected hitch. Claudia Marx was understanding as a fellow-traveller on the PhD voyage. Michael and Annette Farrer very kindly provided an out-of-the-blue gift that solved a pressing logistical problem.

Carys Underdown, H. Markettos and Saar Drimer were most helpful as my proofreaders.

I am also very grateful to the many members of Cambridge MethSoc over the years for keeping me (in)sane through difficult times.

Finally, I would like to thank Dr Simon Moore for his patience and perseverance as my supervisor, and Carys and my parents for the many facets of their support.

Publications

The following publications have resulted from this research:

- A. T. Markettos and S. W. Moore, 'The Frequency Injection Attack on Ring-Oscillator-Based True Random Number Generators'. In *Proceedings of Cryptographic Hardware and Embedded Systems (CHES) 2009*, Lausanne, Switzerland, September 2009. Lecture Notes in Computer Science 5747, Springer, pp. 317-331.
- H. Li, A. T. Markettos and S. W. Moore, 'A Security Evaluation Methodology for Smart Cards Against Electromagnetic Analysis'. In *Proceedings of the 39th IEEE International Carnahan Conference on Security Technology (ICCST 2005)*, Las Palmas de Gran Canaria, Spain, October 2005, pp. 208-211.
- H. Li, A. T. Markettos and S. W. Moore, 'Security Evaluation Against Electromagnetic Analysis at Design Time'. In *Proceedings of Cryptographic Hardware and Embedded Systems (CHES) 2005*, Edinburgh, UK, September 2005. Lecture Notes in Computer Science 3659, Springer, pp. 280-292.
- A. T. Markettos and S. W. Moore, 'Electromagnetic Analysis of Synchronous and Asynchronous Circuits using Hard Disc Heads', 16th UK Asynchronous Forum, Manchester, UK, September 2004.

CONTENTS

Chapter 1
Introduction

Identity and *integrity* are vital parts of many communication systems. Each party needs to verify the authenticity of the other, and that the communication has not been modified in transit. In banking an error in either allows theft of money.

In security identity is provided by use of a *key*, a secret known only by one or both parties. Using cryptography they can demonstrate possession of the key to assert their identity. Since the key is simply a piece of information, it can be easily copied. To avoid this, the key is often held inside a *hardware security module* (HSM) which performs the cryptography. The key is deeply embedded in the module so that it is technically difficult and expensive to extract it.

The *smartcard* is a low-cost HSM. Costing $1–10, a silicon chip will store the key and allow only chosen cryptographic functions to operate on it. These have been widely distributed as banking cards and Subscriber Identity Modules (SIMs) for mobile phones. An attacker must take apart the chip to discover the key, which is difficult and costly.

An alternative approach is to infer the key from the chip's behaviour. *Power Analysis* records the power consumption of the cryptography, in the hope of deducing the key. *Electromagnetic Analysis* (EMA) instead uses the electromagnetic emissions of the device.

Performing an EMA attack is practically more complex than power analysis. It is much more frequency-dependent, and the receiver design depends on the frequencies of interest. Awareness of the frequency domain and high frequency measurement techniques are important for the attacker. An attack requires much heavier tailoring to the application, or expensive equipment in the absence of targeted hardware design. It is thus more difficult to generalise equipment over many targets.

Such attacks are *passive attacks*. They simply involve listening to the effect on the environment caused by the secure circuit. *Active attacks* also exist, tampering with the operation of the circuit. An example is *glitch attacks*, where signal or power inputs are changed in ways that violate the device's specified conditions for correct operation, usually by applying fast transients. Another example is the optical fault attack, where specific transistors are targeted with a laser. The aim is to include a fault in the cryptography that leaks key material.

1.1 MOTIVATION AND CONTRIBUTION

In this thesis, I expand EMA into the field of active attacks. In particular I aimed to investigate the feasibility of frequency-based active attacks. Electromagnetic fields are ubiquitous in the universe. Since the advent of radio, targeted broadcasting of high powered electromagnetic radiation has been used around the world. A powerful broadcast station such as the Taldom longwave transmitter near Moscow emits 2.5 MW (Bobbett 1999, p. 276). Any electronic device must operate in the presence of these background signals, and so must have a certain degree of immunity to them.

On the other hand, electromagnetic effects can be very subtle and operate on a large dynamic range. The signals from a deep space probe, such as the Galileo probe to Jupiter, are received on Earth at a power of around ten zeptowatts, i.e. 10^{-20} W (National Aeronautics and Space Administration 1996). Yet with suitable antenna(s), receiver and signal conditioning they can be successfully decoded.

Thus a secure circuit may be tricked into radiating secret information by interactions of the frequencies generated both inside and outside, and this may be difficult to suppress. Given pre-existing electromagnetic interference (EMI), it may already being doing so if anyone troubles to look.

In Chapter 2 I introduce the field in greater depth and examine the literature. In particular I discover that such effects are already known by the military community, though much of this work remains classified.

Chapter 3 describes the theoretical background and some of my experimental apparatus. EMA is a field where the choice of equipment matters a great deal more than that used for power analysis. In particular I describe some of the techniques necessary to achieve useful results. I also outline the design of a test chip I had fabricated to evaluate the electromagnetic properties of recent semiconductor processes which I used for later measurements. A more detailed data sheet may be found in Appendix A.

Chapter 4 then evaluates a range of sensors for measuring the electromagnetic properties of integrated circuits (ICs). They are examined for their sensitivity, their frequency response, and their spatial resolution. This provides a toolkit with which I can make electromagnetic measurements.

Chapter 5 examines the behaviour of an integrated circuit when it is exposed to a frequency, in this case by injecting it into a power supply. I examine the various modes by which features of its internal operation may be modulated by signals added by an attacker, and how such modulation can be exploited to aid the attacker. I demonstrate this with a timing attack

on a 'secure' microcontroller and how it may be tuned to an attacker's receiver.

Finally Chapter 6 investigates altering the behaviour of a secure circuit by means of frequency injection. I cause the random number generators in a number of secure circuits to malfunction and give poor randomness. Poor randomness is the cause of many vulnerabilities in cryptographic protocols. I demonstrate the flaws in two commercial devices: a commercial banking card and a microcontroller used in Automatic Teller Machines (ATMs). I also describe a realistic scenario where theft might be possible using this vulnerability, and how it may be prevented.

CHAPTER 2
BACKGROUND AND OVERVIEW

> *"A little reflection," continued Frank, "soon convinces a man that rough downright stealing is an awkward, foolish trade; and it therefore falls into the hands of those who want education for the higher efforts of dishonesty. To get into a bank at midnight and steal what little there may be in the till, or even an armful of bank-notes, with the probability of a policeman catching you as you creep out of the chimney and through a hole, is clumsy work; but to walk in amidst the smiles and bows of admiring managers and draw out money over the counter by thousands and tens of thousands, which you have never put in and which you can never repay, and which, when all is done, you have only borrowed;–that is a great feat."*
>
> ANTHONY TROLLOPE, *The Eustace Diamonds*, 1872

2.1 BANKING SECURITY

Money, it is said, is the root of all evil. While that may not literally be true, there are many who wish to transfer it from those whom society holds to be its legitimate owners.

It is no longer necessary to climb down the chimney into the strong-room and escape with the gold bars. Indeed, today there are no strong-rooms and no gold bars. Money is no longer held in physical objects, but in the computers of governments and banks around the world. Central banks can 'print' money by merely altering the programming of their computers.

Moreover, at the same time as the number of transactions has increased, the number of staff overseeing them has decreased. Banking is no longer a face-to-face transaction conducted with those you know, it is with a computer to whom you are just a number. Banking transactions are simply flows of information between computers.

Thus the security problem is no longer one of vaults, locks and guards, but one of information security. The modern bank robber does not use a shotgun, he uses a computer.

In a world where information is virtually free to store and to transmit, the crime of impersonation is all the more easy to commit. The genuine customer no longer visits a physical building to make a transaction, they

do it by telephone or by an exchange between computers. With no means to pass physical objects, the bank must determine if they are an impostor by checking if they know certain secret information. But if the impostor is listening, he can learn it too.

One defence is the use of cryptography. The bank and customer can share their secrets over a public channel, such that nobody else can understand them. But this channel is secured by yet more pieces of information. If the impostor discovers these, he can once again impersonate the customer or the bank. Or he can impersonate both, pretending to be the bank to the customer and the customer to the bank.

Another defence is to revert to using physical objects or tokens. The customer can hold a particular physical object and the bank can ask questions of it. If the questions are difficult but answered correctly, it is quite likely that the real physical object is being held. This requires an object that is unique, difficult to copy, but, at the same time, cheap to mass produce.

Here the silicon integrated circuit comes to the rescue. It is cheap, easy to mass produce, and able to perform cryptographic operations that are rigorous enough to minimise risks of compromise by a flaw in the mathematics. However uniqueness is the problem. The whole semiconductor industry has put 50 years' worth of effort into eliminating variation in its manufacturing processes.

The uniqueness problem may be solved by giving the device a key, an identifier that is different for each one, but this reduces the work of the impostor. He can build an accurate simulator of the rest, and only need worry about extracting the key. He is especially interested if the same key is used in multiple devices, since he can destroy one to extract the key, then use that key to impersonate another.

Since the 1970s the field of tamper-proof hardware has developed. Here the keys are held in storage deemed to be very difficult to attack, either due to its small size on an integrated circuit, or with active protection that wipes the keys when an attack is detected.

The cheapest tamper-proof hardware device is the smartcard. Here a small and cheap processor is attached to a plastic card, with the keys held deep within the processor. These are used as security tokens for banking, and their low price has expanded their use into other areas which require verification of entitlement to goods (e.g. meals, cigarettes), services (e.g. transport tickets, telephone calls) or rights (e.g. voting in elections, crossing international borders).

But how good is their protection? This depends on the design of the smartcard and its implementation. Much work has been done on evaluating attacks aimed at extracting the keys and the defences that foil such

attacks. This 'arms race' is set to continue. Only by investigating likely vulnerabilities can we discover their effectiveness – or learn if they are already in use by others for good or ill.

2.2 THE ORIGINS OF THE SMARTCARD: TOKENS FOR CREDIT

The concept of money has existed since ancient times. Banking was well established in the Greek and Roman empires, but fell out of use in the West until the Middle Ages. Banknotes, however, were invented by the Chinese but only introduced in Europe in the sixteenth century.

The history of credit goes back into antiquity. In the Greek and Roman Eempires bankers provided finance to merchants carrying freight by sea and for the construction of public buildings. But consumer credit, until the late nineteenth century, was only available from informal sources. Consumer culture, the expectation that 'good living' depends on owning lots of goods, started at this time and expanded in the twentieth century. Consumer culture was fuelled by the expansion of consumer credit. It is to the expansion of credit that we should look to discover the origins and development of credit tokens.

The concept of a token which is used to give access to credit has existed for almost a century. A metal dog-tag was introduced by Western Union in the USA in 1914 (MacDonald and Gastmann 2004, p. 227). The tag identified the holder and permitted them to run up an account to be settled at the end of the month. Credit was now available without a personal relationship between the borrower and the outlet.

The cardboard Diners Club card was introduced in the USA in 1950 (Rankl and Effing 2004, p. 2). It rapidly gained acceptance as a status symbol, and was the first credit card to achieve widespread use.

Such cards depended on the honesty of the merchants and holders. As the cards became more common, fraud entailed some form of duplication of the card. The Eurocheque card (1968) introduced security features using a watermarked security paper laminated between polyvinyl chloride (PVC) layers (Haghiri and Tarantino 2002, p. 20).

Earlier, in 1960, Bank of America had introduced plastic credit cards. These were made of a single layer of PVC with the customer's name and account number embossed in gold or in a contrasting colour. At the time colour printing was generally only available using rotary printing machines which could not print on the semi-rigid plastic cards. Thus only specially designed manufacturing plants could produce the cards. As the

cards were rare and considered valuable, vendors carefully checked the signature (Hendry 2001, p. 35).

Today printers for printing bank-sized cards can be bought for under $1000 (PC World Business 2010). As the technology for duplicating cards improved, banks had to keep ahead. Further security features were introduced, such as different types of security printing to defeat naïve photocopying. In parallel with such developments, the technology was applied to tokens used in exchange for services, such as transport tickets or telephone cards, not just directly for cash.

Any printed security features depend on the sales assistant having the time and skills to check them. Holograms today provide a quick check that the card belongs to one of the two main payment schemes: Visa or Mastercard (Hendry 2001, p. 35).

The magnetic stripe card was invented by Forrest Parry at IBM in the early 1960s (Southern Utah University 2004, p. 24). It was initially deployed at the Central Intelligence Agency, then for transport systems such as the San Francisco BART, and then added to credit cards. The magstripe was the first card that was machine-readable; all previous cards had relied on human verification.

The idea of an integrated circuit on a personal card was proposed by Dethloff and Grotrupp in 1968 (Rankl and Effing 2004, p. 3), and patented in 1970 (Dethloff and Gröttrup 1972). On the card was an integrated circuit containing the card's identifier (secret) which would be difficult to copy. This was communicated via a radio frequency, capacitive, optical or electrical link between the terminal and the card. Guillou (2004) shows the relationship of the many pioneering patents at around this time.

Ellingboe (1972) describes a similar device in his patent (which was originally applied for in 1967). Here a 'Data Processor' is connected via a serial interface to a series of pads on the edge of the card.

Halpern (1975) describes a card with a fraud counter: if the externally supplied validation character does not match the internally held value, the counter is incremented. After six occurrences, the card permanently disables itself.

Moreno's patent (Moreno 1975) was the first to be used commercially. Silicon integrated circuits had advanced sufficiently that the functions might be realisable on a single chip. CII Honeywell Bull acquired a licence in 1974 (Guillou 2004).

In 1977 Michel Ugon from CII filed a patent for a smartcard including a microprocessor (Ugon 1980). This incorporates a separate key-holding memory inaccessible from outside.

The first functional device in the smartcard format was a proof-of-

concept, produced in cooperation with Motorola and released on 21 March 1979 (Guillou 2004). Called the CP8 (Girardot 1984)[1], it contained a Fairchild 3870 CPU and 2716 EPROM. It was used in an experimental telebanking system with La Poste.

There were several other early prototypes, which are described in Guillou's (2004) history of smartcards.

In 1984 the French postal and telecommunications services agency, PTT, carried out a successful field trial with telephone cards. Germany carried out a similar pilot in 1984-5, using magnetic stripe, holographic and smart cards in a comparative test. Smartcards provided reliability, security against manipulation, and flexibility for future applications. Both France and Germany rolled out (mutually incompatible) telephone card systems in the mid-late 1980s. These used EPROM (France) or EEPROM (Germany) on memory-based cards (Rankl and Effing 2004). By 1990, 60 million cards were in circulation in France.

The German Post Office introduced a microprocessor card with EEPROM as an access control to its C-Netz analogue mobile telephone network as magnetic stripe fraud was becoming a problem. This experience led to their use in the successor GSM digital mobile telephone network, which became operational in 1991 (Rankl and Effing 2004).

With growing interest in smartcards, the International Standards Organisation (ISO) decided, in October 1981, to standardise the smartcard format and interface (Guillou 2004). This resulted in the ISO 7816 family of standards in 1987 and 1995, which cover their physical and environmental characteristics (including activity under electromagnetic, X-ray and other stresses).

Building on ISO 7816, the EMV specification (named after its originators Europay, MasterCard and Visa) was first published in 1995 (Messmer 1995). The specification (EMVCo, LLC 2008) allows various modes of operation, but fundamental to it is that the smartcard contains secret keys which are used in payment operations. Instead of a magnetic stripe, where the complete data is returned when it is read, the EMV smartcard performs cryptography to sign banking transactions, so that the bank has confidence that a real card is being used instead of a counterfeit.

[1]CP8 is subsequently used to refer to Bull's smartcard division (Bull Worldwide Information Systems 1996)

2.3 CRYPTOGRAPHY BACKGROUND

Once we have designed a secure tamperproof circuit, how may we use it to protect an activity such as a telephone call or banking transaction?

On the subject of a military cryptography system, Kerckhoff's second Principle states:

> 2. Il faut qu'il n'exige pas le secret, et qu'il puisse sans inconvénient tomber entre les mains de l'ennemi
>
> (Kerckhoffs 1883)

that is, "[the system] does not require secrecy, and may without inconvenience fall into the hands of the enemy".

If the system can be freely given to the enemy, security must rest entirely in the one fact the enemy does not know: the value of the key.

Until the 1930s, cryptography used essentially pen-and-paper systems that could be broken with a greater or lesser degree of statistical analysis and guesswork (the cryptanalysts only having the tools of deduction, pen-and-paper and statistical tables). Herbert Yardley denounced these in his book as 'sixteenth-century codes', with some justification (Yardley 1931, p. 257), (Kahn 1996, p. 362). Such systems were typically ad-hoc, with little mathematical rigour. Algebraic analysis of cryptography was touched upon in a number of obscure publications: by Claude Comiers in 1690 (von zur Gathen 2003), by F. J. Buck in 1772 (Buck 1772, von zur Gathen 2004), and a 1926 detective magazine article by Jack Levine (Kahn 1996, p. 405). A general method was only proposed by Lester S. Hill in 1929 (Hill 1929). With the advent of computers such systems became easy to break, but also more complex systems could be devised.

Modern cryptosystems can be divided into two broad categories: *secret key* or *shared key*, where a single key is shared between both parties and must be kept secret to ensure security and *public key*, where each user has two keys, a private key that they must keep secret and a public key they can broadcast. The public key may be used to encrypt messages to the user, but cannot be used to decrypt their messages. RSA (Rivest, Shamir, and Adleman 1978) is an example of public key cryptography, while the Data Encryption Standard (DES) and Advanced Encryption Standard (AES) use secret keys.

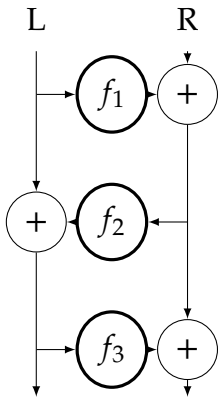


Figure 2.1: *Structure of a Feistel cipher*

2.3.1 The Data Encryption Standard (DES)

The *Data Encryption Standard* (DES) is a secret key cryptosystem that is particularly popular in banking applications. I shall outline it only briefly here based on the description in Anderson (2001, ch. 5): a more detailed description including all the technicalities may be found in Schneier (1996, ch. 12).

DES is a member of the class of ciphers known as block ciphers, that is they process the input in blocks of a fixed size, producing an output in blocks of a fixed size. This means that data must be parcelled up into blocks with padding if necessary, and that care is required so that it is not possible to cut-and-splice blocks of ciphertext to modify the message without knowing the key. This is prevented by 'block chaining' schemes (outlined in Schneier (1996, ch. 9)).

It is based on the *Feistel cipher* named after Horst Feistel and developed in the 1960s. A Feistel cipher has the structure seen in Figure 2.1, based on a number of 'rounds'. Feistel ciphers divide the input into two halves, left and right, and apply them to a ladder structure. In each *round n*, a function f_n is applied to one half of the data, and the output is exclusive-ORed (denoted by $\oplus$, since it is binary addition without carry) with the other half. In the next round, f_{n+1} is applied to the second half of the data, and the result exclusive-ORed with the first half, and so on. If the total number of rounds is even, the left and right halves are swapped at the end.

The main point of a Feistel structure is that decryption is performed using the round functions in reverse order, and thus f_i does not need to be invertible. This means any one-way function may be turned into a block

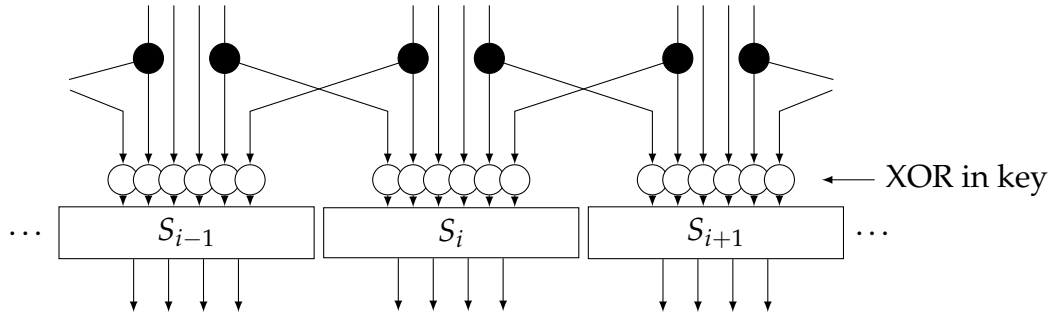


Figure 2.2: *The DES round function*

cipher. If all f_i are the same but for the key, we can use the same hardware to encrypt as to decrypt by reversing the order we supply the key.

DES is a Feistel cipher, using a 64-bit block, a 56-bit key and 16 rounds. Each round operates on two 32-bit half-blocks, and the function f_i consists of:

- Expansion: The half-block is first expanded from 32 to 48 bits. This allows every bit of the plaintext to affect two bits of ciphertext.
- Key mixing: These 48 bits are then exclusive-ORed with 48 bits of round key
- S-box substitution: The result is passed through 8 S-boxes, which each take 6 bits of input and produces a 4 bit output
- Permutation: The 32 bits of output are permuted in a fixed bijective permutation (each input bit has exactly one output bit)

The structure of the expansion, mixing and S-boxes may be seen in Figure 2.2.

The 56-bit key is transformed into 48 bits per round using a compression permutation, selecting a different 48 bits from the key for each round.

The 56-bit key for DES is now regarded as weak: the current record for a brute-force attack on average is one day (SciEngines GmbH 2009). One remedy used by EMV is *triple-DES*, or 3DES, which performs DES three times with independent keys. If a single DES operation is $DES(k; M)$ for key k and message M, Triple-DES is described by:

$$3DES(k_0, k_1, k_2; M) = DES(k_2; DES^{-1}(k_1; DES(k_0; M))) \qquad (2.1)$$

If all three keys are the same, the inner two operations cancel out and 3DES reverts to DES. This allows a backwards compatibility mode. Full

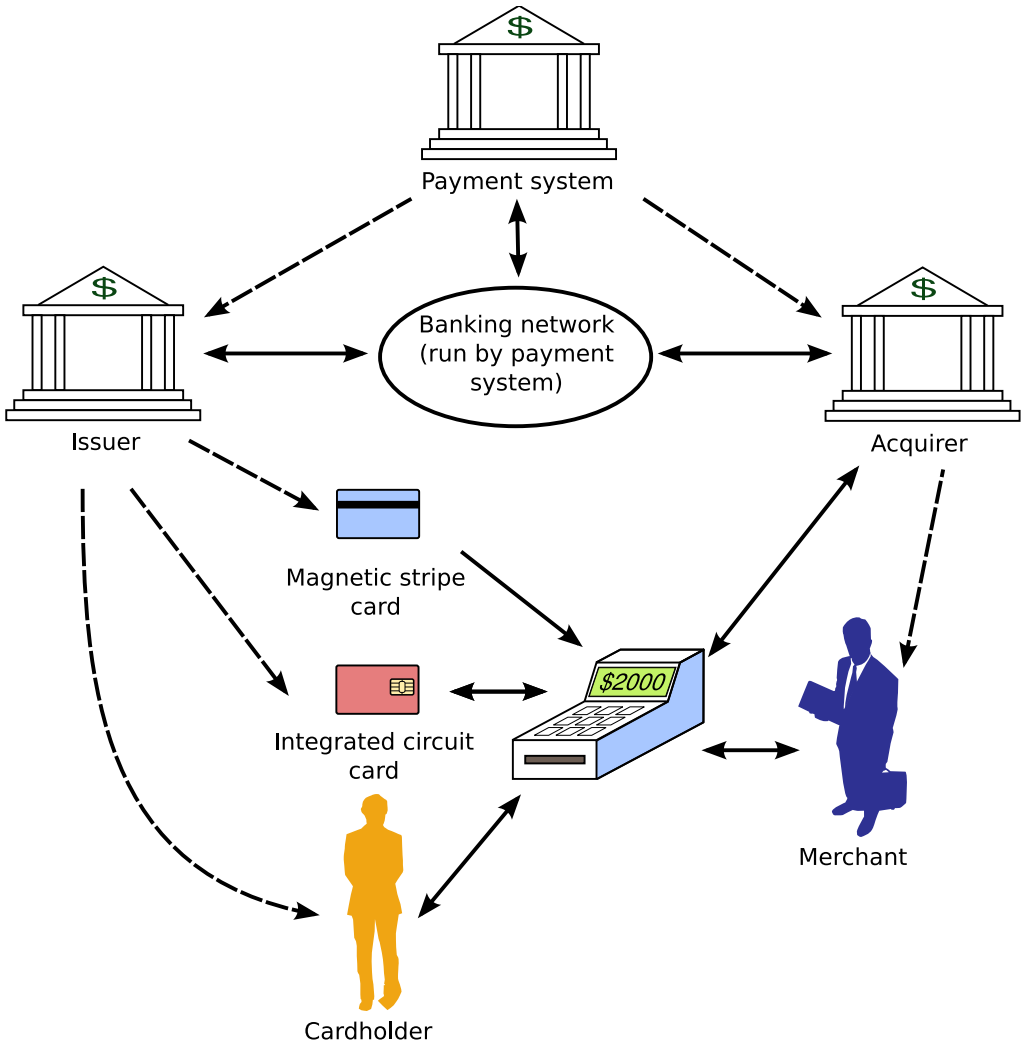


Figure 2.3: *Parties involved in EMV (based on figure 1 from EMVCo, LLC (2007))*

3DES has a key length of 168 bits, well outside the range of brute-force attacks. A reduced form of 3DES exists called *two-key triple-DES*, which sets $k_2 = k_0$, giving a key size of 112 bits.

2.3.2 The EMV protocol

EMV defines a subset of the communications in a complex web of relationships between parties in a payment system (Figure 2.3).

Of interest here is the part of EMV concerned with communication

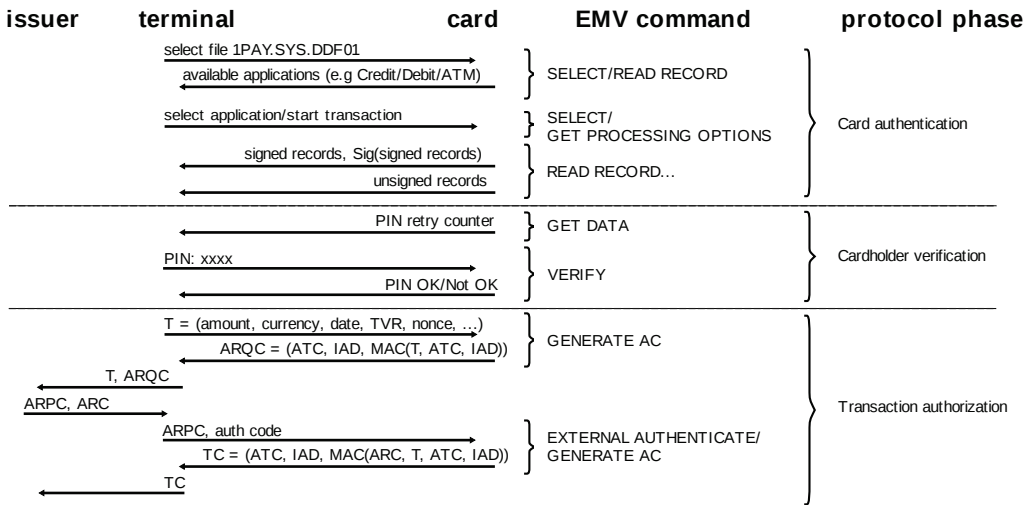


Figure 2.4: *An EMV transaction (reproduced from Murdoch, Drimer, Anderson, and Bond (2010))*

between the Integrated Circuit Card (ICC) and the terminal. An EMV transaction can be seen in Figure 2.4.

An EMV transaction consists of three phases, namely card authentication (checking that the card has been issued by a valid bank and has not been altered), cardholder verification (checking that the card is being offered by someone authorised to use it, usually by a PIN or a signature) and transaction authorisation, where the card is used to authorise this particular transaction.

EMV has three possible authentication methods, which a terminal may perform online (with a connection to the bank) or offline (when the transaction is of low value, the network is down, or a connection is infeasible such as on an aircraft):

- Static Data Authentication (SDA): the terminal verifies the signature of static data on the card, data written by the bank and signed at card personalisation before shipping to the cardholder. The card only requires symmetric cryptographic hardware, which makes it cheaper.
- Dynamic Data Authentication (DDA): the card signs the transaction data with its private key, to ensure that the card is not counterfeit and the card data has not been altered. This requires public key cryptography hardware on the card, which is more expensive.
- Combined DDA/Application Cryptogram Generation (CDA): the card signs the transaction data plus the online cryptogram, to en-

sure the cryptogram has not been modified between the card and the terminal

In offline authentication, the terminal receives a cryptogram (an Application Request Cryptogram, or ARQC) supplied when the card authorises the transaction, but has no means to verify it. In online authentication, the transaction and cryptogram are sent to the issuing bank who returns an Application Response Cryptogram (ARPC) and Authorisation Response Code (ARC), which are passed on to the card. Finally, in online or offline cases, the card returns a transaction cryptogram (TC) to approve the transaction.

EMV uses several cryptographic primitives. For signing records and checking signatures, RSA public key cryptography is used. 3DES is used to generate session keys and for computing Message Authentication Codes (MACs)[2].

The ARQC contains a MAC of various transaction data (amount, currency, country, date, etc), an unpredictable number, a transaction counter, and Issuer Application Data (IAD), whose meaning is only known to the card and the issuer. Similarly the TC contains a MAC of the same data plus the ARC. The MAC is a 3DES encryption of the data, padded in a specific way, using a secret session key generated from the card's master key and the transaction counter.

As we shall see later, many smartcard attacks require repeating transactions with the same key. EMV is well-defended against this as it takes care never to use a key more than once. Counters and unpredictable numbers make the job of the attacker harder by also varying the data.

2.4 SECURITY STANDARDS

Various standards exist for certifying the security of a computer system, typically used by vendors to assert they have reached a certain level of security.

2.4.1 FIPS 140

The Federal Information Processing Standards series 140 (National Institue of Science and Technology 2001) provides various levels of certification and specifies the security properties a cryptographic module must meet to be certified at each level:

[2]described in EMVCo, LLC (2008, vol. 2, sec. A1.2)

- FIPS 140 level 1: Basic functional security requirements, no physical security
- FIPS 140 level 2: Tamper evidence: the need for tamper-evident coatings or seals, or for pick-resistant locks on covers of the module. Rôle-based authentication means the module will only perform operations for an operator authorised to use them.
- FIPS 140 level 3: Tamper detection: the module is able to detect tampering attempts, and zeroes critical security parameters (CSPs) such as keys. Plaintext CSPs must be entered via a separate path from that used for normal operation.
- FIPS 140 level 4: Tamper detection and environmental monitoring: an even stronger likelihood of tamper detection, such that the module is completely surrounded by a tamper detection envelope. The device must also be protected against external variations in voltage or temperature that are applied by an attacker – either to be unaffected by them, or to zero the keys when detected.

The FIPS requirements are mostly physical security requirements: they concern modifying the hardware. This is only one part of the picture. For example, the IBM 4758 PCI Cryptographic Coprocessor is certified at FIPS 140-1 levels 3 and 4. It has a tamperproof membrane and an onboard battery that will zero the keys if tampering is detected. But Bond (2000) managed to extract keys using a flaw in the software API used to access the device. There was no need for complex drilling of the security membrane when the device could simply be commanded to give up its keys.

2.4.2 Common Criteria

The Common Criteria for Information Technology Security Evaluation (Common Criteria or CC for short) are a series of standards aimed at defining security evaluation (see reference ISO/IEC (2009)). They have been ratified as international standard ISO/IEC 15408.

They are a multinational effort to define how a target of evaluation (TOE) may be tested against a defined set of security requirements. The target may be an IT device, a software application, or a networked system. The evaluation aims to validate claims made about the target.

First, the requirements of the application are defined in an implementation-independent way in a Protection Profile (PP). Then the

PP is refined into a document giving a series of product-specific requirements, the Security Target (ST).

The Security Target may be composed of Security Functional Requirements (SFRs). The Common Criteria standards provide a catalogue of such functions, such as anonymity, the creation of a secure audit trail, non-repudiation of receipt, and so on. The standards describe how one requirement might depend on another, and what the target should specify in order to be evaluated against each requirement.

The results of an evaluation are produced in a series of Evaluation Assurance Levels:

- EAL1: Functionally tested
- EAL2: Structurally tested
- EAL3: Methodically tested and checked
- EAL4: Methodically designed, tested and reviewed
- EAL5: Semiformally designed and tested
- EAL6: Semiformally verified design and tested
- EAL7: Formally verified design and tested

With increasing rigour of evaluation comes increasing costs, so the highest levels are typically only used where the risk of failure justifies the costs.

Many smartcards have been certified at EAL 3 or EAL 4, with many PIN entry devices for terminals being certified at EAL 4. As Drimer, Murdoch, and Anderson (2008) showed, the simple fact of being certified is not protection against attacks which were not conceived by the author of the evaluation requirements.

2.5 ATTACKS ON SMARTCARDS AND SECURE CIRCUITS

Having established an electronic token whose correct function is required for access to money or services, how can this be subverted?

Very simple attacks are physical ones: stealing the card, printing a non-functioning duplicate, or coercing a holder to make an unwanted transaction. These are well-understood, and easy to detect.

More technical attacks involve tampering with the device itself. For example, early telephone cards used EEPROM memories. By sticking tape over the EEPROM programming voltage which was supplied on a different pin, it was possible to prevent the number of remaining units from being decremented. A small amount of thought provides simple countermeasures to this attack (decrement the unit count first and check it has taken effect, before allowing the call).

Other attacks are technically simple but require more preparation. You could copy the magnetic stripe from a victim's card onto your own. Such an attack requires knowledge of the victim's PIN, and access for a brief period to their card. The PIN could be determined by threatening them (but then the victim might tell the police) or by installing a camera in a cash machine and a magstripe reader to read their card as it goes in. For that reason modern cash machines draw the card in jerkily, to prevent an even reading of the magstripe.

The forger's ultimate aim is to copy the token. The forger can then do their impersonation without having to be anywhere near the victim (and thus the police). The victim may not even realise that a copy has been made, and so they carry on oblivious. If the token looks robust, when the customer discovers a problem and complains they may not be believed.

Copying magnetic stripe cards is straightforward: the equipment costs a few hundred pounds. The rollout of smartcards was intended to increase this cost substantially. Unlike a magnetic stripe card, there is some data which a smartcard will not reveal. The card is intended to operate as a black-box oracle, without ever revealing its secret keys to the outside world. The card will only perform limited operations with them, using a limited interface.

But smartcards, and other consumer security devices, operate in a hostile environment they cannot control. They are handed out to every bank customer, so the attacker can easily get hold of some samples to take apart in private, and practise their techniques. Furthermore, the cryptography the smartcards perform is only valid if correctly performed. If not correctly performed, at best the result is wrong; at worst some secrets may be leaked. And all the while the attacker can observe the operation, looking for tell-tale clues to the secrets.

2.6 SIDE-CHANNEL ANALYSIS

Any computational system can be defined as a system that takes an input, performs some computation, and produces an output. This output is com-

posed of two components: the main channel – usually the expected output – and a side-channel, consisting of other undesired outputs during and after the computation. While the main channel is typically in a mode designed to be interfaced to the next component, usually electrical or optical, the side-channel may involve energy (and hence information) leakage in many different forms. This leakage may contain information about the computation being performed. In a security context, if the computation being performed is deemed to be secret, the information leaked may compromise that secret. Whilst typical secure algorithms are designed such that the main channel output contains no information that would compromise the secret computation (essentially the secret key), they are usually not designed to protect against side channel emissions. In principle this is because side-channel emissions take diverse forms and are typically implementation-dependent.

Naccache and Tunstall (2000) cite three means of attack via side-channels namely timing, power and fault induction.

Power analysis opens up a wider field of measuring some property of the cryptosystem. It is not only relevant to electrical power, but to any modes of energy transfer into and out of the system, including conducted electrical power, conducted mechanical vibrations (infrasonic, audio, ultrasonic) and electromagnetic radiation (radio, thermal/infra red, optical, gamma/X-rays). In the literature, the phrase 'electromagnetic emanations' typically refers to radio or microwave emissions. Different modes can also be used in combination.

Timing provides another angle, since we can observe variations in timing of energy transfer as further information leakage. As well as straightforward temporal variation of signals, this also applies to the frequency domain.

In addition, any computation system must exist in the physical plane so we must remember to consider the spatial dimensions.

2.7 SIDE-CHANNEL MODES

While this multi-dimensional vector space is useful for placing side channel emissions in a theoretical context, we need, on a more practical level, to analyse side-channel *attacks*, which involve the use of side-channel leakage against a cryptographic system or algorithm.

2.7.1 Timing analysis

Kocher (1996) proposed the idea of a timing attack, being one based on the fact that the time an algorithm takes to complete may depend on the secret key. By timing encryption of different data with the same key, it is possible to make statistical guesses as to the key.

Tsunoo, Saito, Suzaki, Shigeri, and Miyauchi (2003) point out that, with modern computer architectures (pipelines, caches, etc), it is difficult to make all programs take the same length of time, and hence the time taken leaks information about the operation being performed. They demonstrate attacks on DES using timing information leaked via the CPU cache.

2.7.2 Power analysis

Electronic engineers have long been aware that fast signal transitions result in spikes of current being drawn from the power supply. Back in the 1920s (Ballantine 1923, pp 262–269), the coupling of signals to power rails became a problem in valve equipment, and bypass capacitors across the supply rails were suggested to reduce this (Carr 1928, Popular Mechanics 1929). Since the earliest transistor logic, the placement of decoupling capacitors around the circuit, providing localised smoothing of these spikes, has been necessary to avoid undesirable behaviour due to voltage dropouts (Trent 1952).

However, it was a long time before such techniques were used to infer secret information. Kocher, Jaffe, and Jun (1999) were the first to demonstrate how the power consumption of a device could be used to extract the secret keys. Messerges, Dabbish, and Sloan (1999) illustrate this with practical results relating to smartcards, particularly those performing DES.

Simple power analysis involves recording the power consumption of a cryptographic operation and using that to deduce the operations taking place and sometimes the key. Either the absolute power consumption may be different, or its distribution in time. For example, an implementation of modular exponentiation might use a square and multiply algorithm. In this algorithm, exponentiation is achieved by a repeated squaring of the base for each binary bit of the exponent and the result of the square is held in a variable. Where the binary bit of the exponent is a one, the result variable is multiplied by the current value of the square. The power traces of these two operations are different, so, looking at the power trace, the attacker can read off the key bits that are ones (multiplies between squares) and those that are zeros (adjacent squares).

Kocher, Jaffe, and Jun's (1999) main contribution was *Differential Power*

Analysis (DPA). This is a statistical technique aiming to extract key bits from noise. The attacker knows that the device has a fixed key, but not what it is. The key is too long to brute-force. A bruteforce attack is regarded as the baseline cryptographic attack since, in the absence of higher level controls, it is guaranteed to succeed. The attacker tries every possible combination of the key until they land upon the correct one. Typically, key lengths are chosen such that a bruteforce attack takes an infeasibly long time, perhaps millions of years, given current and prospective future technology.

The attacker also knows the encryption algorithm being performed and the details of its operation. He has physical access to the device and is able to make detailed recordings of the power consumption of the device over the duration of the crytographic operation. The utility of DPA is its ability to find out the key in chunks, so a full bruteforce is not required.

To begin, he identifies the first chunk and how many key bits are used. This might be the first or last round of a DES or AES operation. He records many power measurements of the operation for a number of different plaintexts or ciphertexts. He might perform 1000 DES operations and store the ciphertexts in an array C_i where i is the index of the operation (0–999). The power traces are recorded in an array $S_{i,t}$ where t is the index of the time sample recorded by the oscilloscope (perhaps 0–99,999 for 100,000 data samples).

Then, for each ciphertext value C and a single bit b of the intermediate value (perhaps $b = 0..31$) of the computation, the attacker chooses a partition function $D(C, b, K_i)$, taking input of a few key bits K_i. The aim is to select the value of a single-bit result (perhaps the most significant bit) of the sub-computation performed on K_i and C. If K_i is incorrect, the probability of $D(C, b, K_i)$ being correct will be about $\frac{1}{2}$. If K_i is correct, the probability will be 1. $D(C, b, K_i)$ is deterministic: it is purely a model and can be accurately calculated by a simulation.

Using the values of $D(C, b, K_i)$, we then partition both the real recorded traces and the power consumption model traces into two groups: one where $D = 0$ and one where $D = 1$. We then take the mean of each group and subtract the two means:

$$T(t, C, b, K_i) = \frac{1}{|D_0|} \sum_{S_x \in D_0} S_i(t) - \frac{1}{|D_1|} \sum_{S_x \in D_1} S_i(t) \tag{2.2}$$

where $|D_n|$ is the number of members in the set D_n. $T(t)$ is the 'differential trace', a function of the difference of the two means over time (all other C, b and K values being constant for each run). When plotted, the

trace will be roughly flat noise, except if the key hypothesis K_i is correct, which should show as a peak at the time instant b is used. From this, the correct key for the chunk can be determined and thus the whole deduced by unrolling the encryption, chunk by chunk.

An optimisation involves using a model for the power consumption of the circuit to calcuate $D(C, b, K_i)$. This commonly measures the Hamming weight (the number of ones) or the Hamming distance (the number of bits that change state) of the result.

DPA is able to extract keys from traces with much lower signal-to-noise ratios than SPA. If the leaking bit is a single transistor, DPA may still be able to detect it.

There are many refinements of DPA, such as Higher-Order DPA, where multiple samples with a trace are used, or where the traces are partitioned into more than two sets. There is a large body of literature on preventing power analysis attacks, and commercial products from smartcard manufacturers which claim to include security features against power analysis. A collection of papers can be found referenced at ECRYPT Network of Excellence (undated). A good background to the field may be found in Mangard, Oswald, and Popp (2007).

2.7.3 Optical side-channel

Skorobogatov (2009) has demonstrated imaging of direct optical emissions of depackaged chips, using infra-red cameras designed for hobby astronomy. Currently a long integration time (many minutes) is required which reduces the possibility for key extraction from a 2D image. It may be possible to extend this technique with statistical processing in a similar manner to DPA. He also demonstrates avalanche photodiode and photomultiplier detectors with a much shorter acquisition time (picoseconds to milliseconds) at the expense of being an integral over the full spatial area of the detector (thus one tube, plus cooling, is required per pixel).

The non-destructive testing community provide many exotic techniques that are out of reach of the non-specialised laboratory. For example, Soltani, Wysnewski, and Swartz (1999) coat a device with fluorescent dye so that it converts field changes into optical signals, but it is unclear whether this effect is sensitive enough for integrated circuit (IC) usage.

Other non-semiconductor optical side-channels include data leakage from status LEDs of communication equipment (Loughry and Umphress 2002) and reflected light from video displays (Kuhn 2003).

2.7.4 Acoustic side-channel

In 1949, Ryon Page, a member of the NSA staff, heard the sound of the rotor cipher machines at their Arlington Hall cryptocenter and wondered if the plaintext could be recovered from the noises. It turned out that it might, but the idea of acoustical analysis was not developed further for some years (Boak 1973, p. 91). When revisited, it was discovered that highly directional microphones were able to produce good signals from some distance, even down telephone lines – and exactly that type of microphone had been used by eavesdroppers[3].

Preliminary results from Shamir and Tromer (undated) indicate that sounds emitted by a PC running GnuPG cryptography provides enough information to identify cryptographic operations, so the acoustical side-channel is also potentially viable.

2.8 Analysis techniques

Having acquired samples from side-channel(s), it is then necessary to analyse them to infer cryptographic information.

The simplest techniques are those set out by Kocher, Jaffe, and Jun (1999). As described earlier, the idea of simple power analysis involves taking a single trace for an operation and inferring data from it. Differential power analysis is a more powerful technique using the difference between traces taken with different cryptographic material. They may be generalised over other side-channels.

Typically, these attacks require running multiple identical traces with different data. If more traces are available with the same data, then noise can be reduced by signal averaging. However, this is dependent on being able to correlate multiple traces in order to successfully average them.

More complex 'template' attacks are outlined by Chari, Rao, and Rohatgi (2002), involving signal estimation theory. If the attacker has a test device which they can control and which is identical to the real device, they can characterise the signal and noise properties of that device. Therefore an attack on a real device can use this to narrow down on key possibilities that produce traces similar to those recorded in the attack.

This team has gone on to implement multi-channel attacks (Chari, Rao, and Rohatgi 2003). Here they present a theoretical basis for using different side-channels together to complement each other, using the deeper statistical modelling proposed in Chari, Rao, and Rohatgi (2002).

[3]The exact location remains classified SECRET and is redacted from the report.

2.9 FAULT INDUCTION ATTACKS

Another class of attack are so-called active attacks, which involve the introduction of deliberate faults into the cryptographic system. Boneh, DeMillo, and Lipton (2001) demonstrated that relatively few random faults can compromise several public key encryption schemes. Hence inducing faults can be a means to recover key material from the system.

Simple attack techniques involve illuminating the device with radiation to induce faults. This might be infra-red heating (Govindavajhala and Appel 2003), ionising radiation (Shirvani 2001) or other electromagnetic fields. Electromagnetic fields are used by the testing community to simulate interference from physical phenomena (Hsueh, Tsai, and Iyer 1997).

More intelligent attacks involve targeting particular areas of a chip. This usually means knowing the layout of the device or reverse engineering it. Such have been demonstrated using either focused flashguns (Skorobogatov and Anderson 2003) or lasers (Samyde, Skorobogatov, Anderson, and Quisquater 2002).

Fault injection can either be used to simply produce an incorrect output (from which secret data can often be deduced) or it can also be used in combination with side-channels. Samyde, Skorobogatov, Anderson, and Quisquater also use a coil to induce currents in a particular point on the surface of the chip. By measuring the effect on the microcontroller supply current, they can gently stimulate the chip, enough to read out the memory contents. Schmidt and Hutter (2007) and Schmidt (2008) apply a coarser approach: by generating sparks above a microcontroller he is able to inject faults. Neve, Peeters, Samyde, and Quisquater (2003) gives a survey of approaches to fault injection attacks on memories.

2.10 HIGHLY INVASIVE ATTACKS

As well as invasive attacks, such as optical probing, which typically involve depackaging a device (Anderson and Kuhn 1996), there is another class of attacks from the semiconductor manufacturing community. Typically, these attacks are used for non-destructive testing (NDT), evaluation or repairs during the manufacturing process. Such processes may often involve thousands or millions of pounds' worth of equipment and a very close proximity to the die – of the order of micrometres.

The most commonly used is the Focused Ion Beam, or FIB. This is able to cut metal traces on dice and lay new ones. However, practical considerations restrict the extent to which large scale use of FIBs is feasible; in

particular, the probability of failure is high.

Other microscopy techniques exist which are able to monitor signals from the die without physical contact. These are mostly based on the Atomic Force Microscopy (AFM) principle. A micromachined cantilever is scanned across the die surface. The cantilever bends due to the electrical, magnetic or topographic properties of the surface, and an electric field must be applied to restore it to equilibrium. The magnitude of this field mirrors the property being measured, and so a map of the surface is built up.

This leads to techniques such as Magnetic Force Microscopy (Hartmann 1999), where the magnetic field of a substrate is measured. Typically this is used for recording media (Aso, Sato, and Ishibashi 1999), and its application to operating circuits is not clear.

Another magnetic technique is Scanning SQUID Microscopy (Kirtley and John P. Wikswo 1999), which is used to measure currents in integrated and printed circuits, mainly for finding defects. Depending on the sensitivity, this may be useful for current sensing for security.

Electrostatic Force Sampling (EFS) involves capacitive coupling of the probe to the signal under test. In this manner Bridges, Noruttun, Said, Thomson, Lam, and Qi (1998) managed to extract a 3 GHz signal from a chip with only 1 fF of extra load capacitance. Another study by Weng, Falkingham, Bridges, and Thomson (2002) achieves quantitative voltage measurement to within 30 mV of 250 Mbit/s signals using this technique.

Electron Beam Testing (Thong 2004) is an extension of the scanning electron microscope, where a point on the die is illuminated with a beam of electrons. The spectrum of secondary electrons emitted from the surface depends on the voltage present at that surface. The bandwidth is low (at 0.1 V, approximately 1 MHz), but strobed sampling techniques can achieve measurements for periods down to 30 to 40 ps. EBT is severely hampered when measuring signals buried below the die surface, particularly those below other traces.

2.11 ELECTROMAGNETIC ATTACKS

In its broadest sense, electromagnetic analysis (EMA) covers all forms of electromagnetic radiation, including radio, microwave, infra-red (thermal), optical and higher frequency radiation. In conventional usage, the term is used in relation to lower frequency emissions, notably in the radio and microwave bands.

Unintentional electromagnetic emissions were mostly considered in

the electrical engineering community in relation to electromagnetic interference (EMI) and electromagnetic compatibility (EMC), in recognition of the fact that electrical circuits emit electromagnetic radiation when operating. This may affect sensitive equipment nearby, particularly radio and TV receivers, but also other vulnerable devices, notably safety-critical ones, such as life support systems. Electromagnetic compatibility demanded by numerous standards (such as IEC61000-4 or the FCC EMC regulations) and regulatory regimes (such as the Electromagnetic Compatibility Directive for products sold in the European Union) specifies the amount of EM radiation a device may emit or must function under.

2.11.1 Military

Ever since the invention of the electric telegraph, governments have been worried about the problem of interception. In the First World War it was discovered that telegraph traffic to British forward trench positions was intercepted by the enemy, who detected induction of buzzer frequencies on the earth return current of the single-wire telegraph system. In 1915 the Fullerphone (War Office 1923) was developed to prevent interception by low-pass filtering the keyed oscillator signal and transmitting a demodulated baseband signal. This made reception by induction much more difficult. The Fullerphone became standard issue to military personnel operating on front lines (Meulstee undated) throughout the Second World War.

Herbert Yardley was head of the American Black Chamber which cryptanalysed diplomatic traffic from 1917 to 1929. He mentions a British report (Yardley 1931, p.17) which stated that, in 1917, German submarines would lie on the bottom of the Atlantic. They would lay wires of several hundred feet in parallel with telegraph cables on the seabed. Operators on the submarines would then transcribe the cable traffic heard by induction.

At this stage cryptography was performed by hand, using codes or ciphers. A diplomat would write a message in cleartext and pass it to a code clerk. The clerk operating the Morse key for transmission would be provided with a paper copy of the ciphertext from the code clerk. There was no other channel between the encryption and the transmission. Except for errors in encryption or transmission, this separation ensured that a side-channel of the plaintext could not seep through into the transmission. The challenge of emissions security began when machines started being used for encryption and decryption, where such separation was not present.

Recently declassified NSA documents (National Security Agency 1972, Boak 1973) indicate that awareness of this problem began earlier than was previously thought. In 1943 backbone communications of the US Army and US Navy were provided by secure teletype, using one-time tapes and a rotor key generator called SIGTOT. Bell Telephone sold the military a mixing device, called 131-B2, which encrypted by combining a tape or a SIGTOT key with a plaintext. One day, in the laboratory, they noticed that a spike would appear on an oscilloscope in a distant part of the laboratory when a mixer was in operation. By studying the spikes, they were able to read off the plaintext as it went through the machine.

Bell Telephone reported this to the Signal Corps, but the Signal Corps did not believe these tiny pips could be exploited in field conditions. So Bell engineers were placed in a building 80 feet across the street from the Signal Corps' Varick Street cryptocentre in New York. They recorded signals for about an hour, and three or four hours later were able to recover about 75% of the plaintext.

In 1951 the CIA rediscovered the problem when playing with the same 131-B2 mixer. Partly due to the compartmentalisation of intelligence work, this repetition of ignorance and reinvention prevailed in the field for the next forty years.

Peter Wright was taken on as Principal Scientist of MI5 in 1951. It was becoming increasingly difficult to run agents behind the Iron Curtain, so he was recruited to provide the technical and scientific means to gain intelligence of the intentions of the Soviet Union for the Security Service. In his memoirs (Wright 1987), he describes a number of schemes related to emissions security.

The first scheme (Wright 1987, p. 19) was a device discovered in the American Ambassador's office in Moscow, in 1951. A routine sweep, using a tunable signal generator to cause a feedback effect, detected a device lodged in the Great Seal of the United States behind the Ambassador's desk. A small wooden box was found with an aerial on top which fed into a cavity. Inside was a metal mushroom with a flat top which could be adjusted to give a variable capacitance, behind which was a gossamer-thin diaphragm. Found to resonate at 800MHz, the device would modulate speech onto an imposed radio carrier, which could then be received a few hundred metres away by the KGB. By 1953 Wright had produced a copy of this device, codenamed SATYR, which was used throughout the 1950s for covert listening.

Clearly the Soviets had a good knowledge of the properties of electromagnetic security. In 1954 they published a comprehensive set of standards for the suppression of radio frequency interference. These were much

more stringent for teletypes and communications equipment than heavy machinery, despite the latter being stronger emitters. (National Security Agency 1972, p. 27)

A device known as 'Special Facilities' used the same principle of modulating an imposed carrier to tap a telephone (Wright 1987, p. 47). Such a device was used to listen to the setting of cipher machines in the Egyptian Embassy in London during the Suez Crisis, which enabled MI5 to break most of their communications.

In 1958 MI5 suspected radios used by their Watchers (agents who tailed foreign diplomats through London) were being monitored by the KGB (Wright 1987, p. 90). Wright designed the system RAFTER to detect the local oscillator in a superheterodyne receiver and thus infer which frequency it was listening on. It detected a receiver in the Soviet Consulate, and this technique was employed to detect covert receivers used by KGB agents until the end of the 1960s.

After the Soviets and the Egyptians, the French Embassy in London was a prime target for MI5 during the negotiations following the British application to join the European Economic Community, in the years 1960-3 (Wright 1987, p. 109). The embassy used two ciphers: low and high grade, generated independently. There were no telephones available to be tapped in the cipher room, so MI5 attempted to break the low-grade cipher by placing a broadband radio frequency tap on the telex cable carrying the cipher, wired to a room in the nearby Hyde Park Hotel. The principle of operation STOCKADE was to detect echoes of the cleartext which might leak through the cipher machine. GCHQ provided a direct tap of the ciphertext to ensure the RF tap was correct. After matching the low-grade ciphertext, the RF tap was connected to a teleprinter, and the cleartext message could be picked out between the ciphertext. To their surprise, another signal was detected, being cleartext of the high-grade cipher conveyed through a plasterboard wall. Further amplification meant the high-grade cleartext could also be read straight off.

2.11.2 TEMPEST and Compromising Emanations attacks

The electromagnetic side-channel became a persistent worry, particularly in embassies and installations on unfriendly soil where a safe distance could not be maintained between cipher equipment and possible eavesdroppers. It was used by the Germans in the First World War, the US in the Second, the Soviets in the 1950s, the British in 1960s, and the Japan-

ese in 1962. Thus all major intelligence agencies independently became aware of it. The effect came to be known as TEMPEST[4] in the Western intelligence community. Latterly they are referred to as 'EmSec' or 'compromising emanations' attacks.

The basic countermeasures suggested by Bell Labs back in the 1940s are still valid, and consist of:

> **Shielding** for radiation through space and magnetic fields;
>
> **Filtering** for conducted signals on power lines, signal lines, etc; and
>
> **Masking** for either space-radiated or conducted signals, but mostly for space.
>
> (Boak 1973, p. 90)

In addition, spatial containment can be used. Since far-field emission power decays with a $1/r^2$ relation to distance r, a suitable (e.g. 30 m) distance around the machine could be secured from eavesdropping devices.

However, these methods only attempt to remove dangerous frequencies, rather than reducing them at source. A 1956 US Navy keying machine used low voltages – one or two volts rather than 60 or 120 – thus reducing the power for emission. They also instituted RED/BLACK separation, where separate RED circuits carry classified plaintext and BLACK circuits everything else (Boak 1973, p. 94).

TEMPEST-resistance was consolidated into a series of US standards, notably, from 1970 NACSEM 5100 (Boak 1973, p. 99) and later NACSIM 5100 (still classified).

2.11.2.1 TEAPOT, NONSTOP and HIJACK attacks

Various TEMPEST documents also mention three other phenomena: TEAPOT, NONSTOP and HIJACK.

TEAPOT is defined as "the investigation, study, and control of intentional compromising emanations (i.e. , those that are hostilely induced or provoked) from telecommunications and automated information systems equipment" (National Security Agency 199x).

[4]TEMPEST is not an acronym, though some have retrospectively fitted letters to it. Intelligence agencies use randomly selected codewords to avoid giving anything away about the nature of the operation.

The definitions of HIJACK and NONSTOP remain classified, despite the publication of a heavily redacted NACSEM-5112 evaluation manual for NONSTOP (National Security Agency 1975),

Anderson and Kuhn (1999) and Kelsey, Schneier, Wagner, and Hall (2000) speculate that HIJACK refers to crosstalk causing modulation of conducted-field signals (for example, a recording on a tape player), while NONSTOP refers to cryptographic hardware being illuminated by a nearby radio transmitter and rebroadcasting modulation of that transmitter.

Indeed, a 1995 TEMPEST document (National Security Telecommunications and Information Systems Security 1995) refers to NONSTOP as being the primary TEMPEST vulnerability of ships. Observing spatial containment is straightforward except in port, but equipment must operate together in a confined space. Cross-modulation is a concern as the signals can be received a long distance away.

AFSSM-7011 (United States Air Force 1998) is more explicit, giving an example of a NONSTOP countermeasure: "separate the radio 20 meters from all RED processors". HIJACK countermeasures given include separating power and signal lines between RED and BLACK equipment. This lends support to the speculated definition.

2.11.3 Smartcards

The electromagnetic side-channel for smartcards was first demonstrated by Quisquater and Samyde in 2001 (Quisquater and Samyde 2001). They constructed a screened smartcard reader with three coils, of unspecified size except the "total diameters of the coils do not exceed 2 centimeters". Their paper indicates several sensors were tested and signal processing was used to reduce noise to produce a differential EMA measurement. They produce 2D maps of the EM emissions of a device, and also state "all treatments for DPA are possible with EMA since this analysis includes at least the same information". They briefly note some potential countermeasures for EMA.

The first public attacks on cryptographic hardware were published by a team from Gemplus (Gandolfi, Mourtel, and Olivier 2001). They found the best results were achieved using solenoids of coiled copper wire of outer diameter 150 to 500 µm. They noted that enclosing the system in a screened Faraday cage had little effect on the measurements, meaning it is possible to perform EMA in a noisy environment. They describe successful DPA and DEMA attacks on three chips implementing 'alleged

COMP128', DES and modular exponentiation respectively. They conclude that DEMA is more powerful than DPA but SEMA is not necessarily more powerful than SPA.

The EM mode is analysed more systematically by Agrawal, Archambeault, Rao, and Rohatgi (2002a). They identify two categories of EM emanations. The first is 'direct emanations', which follow from intended current flows such as the high frequency components of switching transitions. The second is 'unintentional emanations', i.e. modulated signals (amplitude or phase) which arise from coupling between circuit elements. They note the most effective near-field probes is a small plate of copper or silver connected to a coaxial cable, while in the far field they use biconcical and log-periodic wideband antennas and narrowband high-gain Yagi antennas. They also use current probes.

They discover that the EM side-channel provides complementary information to the power side-channel, and that different EM sensors may reveal different compromising information. They go on to discuss means by which software is vulnerable to EMA.

Agrawal, Archambeault, Rao, and Rohatgi (2002b), goes on to describe statistical techniques for analysis of emanations.

Recently Sauvage, Guilley, and Mathieu (2009) have published cartography of an FPGA with a magnetic field probe which is scanned over the chip surface. They also target a decoupling capacitor to measure the power consumption of the whole chip. In Chapter 4 of this thesis I take a similar approach to cartography, but using on-chip ring oscillators in place of an FPGA.

CHAPTER 3
THEORY AND EXPERIMENTAL SETUP

3.1 INTRODUCTION TO ELECTROMAGNETICS

Many security engineers come from a cryptography or computer science background and may be unfamiliar with details of the basic physics of electromagnetism. In this section I present a brief introduction to the key quantities and relations.

The fundamentals of electromagnetics can be represented in four simple relations, known as Maxwell's Equations, which I shall explain briefly here. For a concise derivation see Robinson (1973) or, for more background, see Kraus (1991). There are many other books on the subject.

Maxwell's equations make considerable use of vector calculus, particularly the divergence ($\nabla\cdot$) and curl ($\nabla\times$) operators. Bold type denotes a vector quantity.

3.1.1 Electrostatics

The derivation of Maxwell's Equations begins with the work of Carl Friedrich Gauss on electrostatics in 1835. He describes the behaviour of static bodies with stationary electric charges.

q The **electrostatic charge** on a body. The charge is sometimes expressed in units of the charge on one electron $e = -1.602\,177\,3 \times 10^{-19}\,\mathrm{C}$; in units of coulomb (C).

$\mathbf{E}$ The **electric field**, defined so that the force $\mathbf{F}$ on a stationary body of charge q is $\mathbf{F} = q\mathbf{E}$; expressed in units of newtons per coulomb ($\mathrm{N\,C^{-1}}$), or equivalently volts per metre ($\mathrm{V\,m^{-1}}$).

ρ The **total charge density**, the charge within a unit volume. Made up from the **bound charge density** ρ_b, formed from dipoles formed by a rearrangement of bound electric charges in the material, and the **free charge density** ρ_f, made up of charges brought from outside. $\rho = \rho_b + \rho_f$. Expressed in units of coulomb per cubic metre ($\mathrm{C\,m^{-3}}$).

ε_0 The **permittivity of vacuum**, a physical constant defined as $\varepsilon_0 = 8.854\,187\,8 \times 10^{-12}\,\mathrm{N\,m^2\,C^{-2}}$. When the material is not vacuum the permittivity is $\varepsilon = \varepsilon_r\varepsilon_0$ where ε_r is the **relative permittivity**. In an isotropic material ε_r is a positive real constant for the material concerned, while it may be a tensor in an anisotropic material.

S A closed surface.

$d\mathbf{S}$ A small vector element of the closed surface, with direction normal to the surface.

V The volume enclosed by surface S.

dV A differential volume element of the volume V, a small piece of volume formed by making small increments along the three spatial axes (in whichever coordinate system suits the problem under investigation).

Gauss' Law relates that the integral of the electric field over the surface of a body is the sum of the charges inside:

$$\boxed{\oiint_S \mathbf{E} \cdot d\mathbf{S} = \frac{1}{\varepsilon_0} \iiint_V \rho \, dV} \tag{3.1}$$

In differential form this may be written as:

$$\boxed{\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0}} \tag{3.2}$$

3.1.2 Magnetism

The nature of magnetism can be described in different equivalent ways, depending on the environment being described.

B The **magnetic field**, which can have units of weber per square metre ($\mathrm{Wb\,m^{-2}}$) or tesla (T), where $1\,\mathrm{Wb\,m^{-2}} = 1\,\mathrm{T} = 1\,\mathrm{V\,s\,m^{-2}}$ (and many other permutations of units). In the CGS system the unit is the gauss (G), where $1\,\mathrm{G} = 1 \times 10^{-4}\,\mathrm{T}$.

v The instantaneous velocity of a charged particle moving in a magnetic field.

F The vector force felt by the particle in the field.

The Lorentz Force The force felt by the moving particle due to the magnetic field, given by:

$$\mathbf{F} = q(\mathbf{v} \times \mathbf{B}) \tag{3.3}$$

I An electrical current ...

L ...which flows in a wire of length L.

$d\mathbf{L}$ A small vector element of the wire carrying a current I.

$d\mathbf{F}$ The vector force felt on the element $d\mathbf{L}$.

The Motor Equation The force felt by the wire is:

$$\mathbf{F} = (\mathbf{I} \times \mathbf{B})L \tag{3.4}$$

In differential form, this gives the **Motor Equation** for the force felt by a small element of the wire:

$$d\mathbf{F} = (\mathbf{I} \times \mathbf{B})d\mathbf{L} \tag{3.5}$$

Gauss' law for magnetism Due to the nature of a magnetic field caused by changing electrical currents, the field lines form closed loops and have no start or end. This is stated by Gauss' law for magnetism, in integral form:

$$\boxed{\oiint_S \mathbf{B} \cdot d\mathbf{S} = 0} \tag{3.6}$$

and in differential form:

$$\boxed{\nabla \cdot \mathbf{B} = 0} \tag{3.7}$$

3.1.3 Electromagnetic induction

Moving wire When a wire moves with a velocity $\mathbf{v}$ in a magnetic field $\mathbf{B}$, an electric field is generated:

$$\mathbf{E} = \mathbf{v} \times \mathbf{B} \tag{3.8}$$

Ψ_m The **magnetic flux**, a measure of the number of magnetic field lines cutting a surface:

$$\Psi_m = \iint_S \mathbf{B} \cdot d\mathbf{S} \tag{3.9}$$

In units of weber (Wb), 1 Wb = 1 V s

B The magnetic flux density, the amount of flux per unit area on a surface. Measured in weber per square metre, sharing the same unit and symbol as the magnetic field.

$\mathcal{V}$ **Electromotive force (emf)**. In a closed circuit loop an electric field $\mathbf{E}_e$ may arise from an energy source, such as a battery, in addition to static charges[1]. The electromotive force (in volts) across a section a to b of the loop is the integral of such field around the section:

$$\mathcal{V}_{ab} = \int_a^b \mathbf{E}_e \cdot d\mathbf{L} \tag{3.10}$$

or, over a closed loop:

$$\mathcal{V} = \oint \mathbf{E} \cdot d\mathbf{L} \tag{3.11}$$

Faraday's Law When an open-circuit wire loop is placed in a time-varying field, an electromotive force is felt across the terminals which depends on the changing flux:

$$\mathcal{V} = -\frac{d\Psi}{dt} \tag{3.12}$$

where t is time.

Maxwell's Equation derived from Faraday's Law. These can be manipulated into:

$$\boxed{\mathcal{V} = \oint \mathbf{E} \cdot d\mathbf{L} = -\iint_S \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S}} \tag{3.13}$$

Put into words, in a loop the emf is given by the integral of the emf-producing electric field around the loop, and also by the integral of the time-variation in flux density over the surface of the loop. In differential form:

$$\boxed{\nabla \times \mathbf{E} = -\frac{\partial B}{\partial t}} \tag{3.14}$$

In the general case, the halves from motion and time-change may be

[1]When integrating around a closed loop, static charges cancel, i.e. $\oint \mathbf{E}_c \cdot d\mathbf{L} = 0$

combined into a more general expression:

$$\mathcal{V} = \oint_L (\mathbf{v} \times \mathbf{B}) d\mathbf{L} - \iint_S \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S} \tag{3.15}$$

Motion Time-change

3.1.4 Currents and dielectrics

J **Current density**, the flow of current per unit cross-sectional area of conductor. Expressed in units of ampere per square metre ($\mathrm{A\,m^{-2}}$).

$\mathbf{J_C}$ In a resistor, the conduction current density $\mathbf{J_C}$ is given by $\mathbf{J_C} = \sigma\mathbf{E}$ where σ is the conductivity ($\sigma = 1/R$).

μ_0 The magnetic constant (also known as the permeability of vacuum), defined to be exactly $\mu_0 = 4\pi \times 10^{-7}\,\mathrm{N\,A^{-2}}$.

The permeability is the degree of magnetisation of a material with a linear response to an applied magnetic field. Maxwell's Equations show that permittivity and permeability are related by:

$$\frac{1}{\varepsilon_0 \mu_0} = c^2 \tag{3.16}$$

where c is the speed of light in vacuum.

μ As with permittivity, in a medium the permeability becomes $\mu = \mu_r \mu_0$, being scaled by the relative permeability μ_r. At high frequencies, the losses in a medium (such as ferrite) can be modelled by describing permeability as a complex value with the imaginary component representing the losses (Kazimierczuk 2009, p. 80). Permeability can further be a function of direction, temperature, humidity and other factors.

H The magnetising field, related by $\mathbf{B} = \mu\mathbf{H}$, expressed in units of ampere per metre ($\mathrm{A\,m^{-1}}$). **H** is the modification of **B** in a particular medium. (The relation is more complex in materials with magnetisation which is not proportional to **B**, such as ferromagnetic materials which exhibit hysteresis).

Ampère's circuit law considers a closed loop around a wire in which current is flowing. It equates the integral of the magnetising field

around the loop to the current enclosed by the loop:

$$\oint_L \mathbf{H} \cdot d\mathbf{L} = I \tag{3.17}$$

It is also possible to equate the current to the surface integral of the current density over the loop, thus:

$$\oint_L \mathbf{H} \cdot d\mathbf{L} = \iint_S \mathbf{J} \cdot d\mathbf{S} = I \tag{3.18}$$

P The **polarisation density**: the density of permanent or induced dipole moments in a dielectric material.

D The **electric displacement field**: the field which causes bound charges in a dielectric to separate, defined as:

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P} \tag{3.19}$$

$\mathbf{J_D}$ The **displacement current density**. When a voltage is applied across a resistor, a current flows through it. This is the **conduction current** I_C. When applied to a capacitor, a current flows due to the rearrangement of charges on the capacitor plates. There is no actual current flow through the dielectric, but it appears as if there is. This may be represented by a quantity named the displacement current, I_D. In vector terms, these can be represented in space as current densities, the conduction current density $\mathbf{J_C}$ and the displacement current density $\mathbf{J_D}$. Thus:

$$\mathbf{J_{total}} = \mathbf{J_C} + \mathbf{J_D} \tag{3.20}$$

The displacement current density is given by:

$$\mathbf{J_D} = \frac{\partial}{\partial t}\mathbf{D} = \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} + \frac{\partial \mathbf{P}}{\partial t} \tag{3.21}$$

The first term gives the contribution from changes in the electric field and the second from changes in the polarisation density of the material.

Maxwell's correction to Ampère's law. Considering also the displacement current with Ampère's law, we find:

$$\boxed{\oint_L \mathbf{H} \cdot d\mathbf{L} = \iint_S (\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}) \cdot d\mathbf{S} = I} \tag{3.22}$$

or, in differential form,

$$\boxed{\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}} \tag{3.23}$$

The boxed equations were presented by Maxwell in his paper *On Physical Lines of Force* (1861).

3.1.5 Electromagnetic propagation

The addition of the displacement current was the key step forward proposed by Maxwell. This extends electromagnetics from consideration of charges and currents in a medium to propagation into space. Considering Maxwell's Equations without any supplied charges or currents, it is possible to derive two wave equations:

$$\nabla^2 \mathbf{E} = \mu_0 \varepsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{3.24}$$

$$\nabla^2 \mathbf{B} = \mu_0 \varepsilon_0 \frac{\partial^2 \mathbf{B}}{\partial t^2} \tag{3.25}$$

These are analogous to the *wave equation* in mechanics. They describe how a wave may propagate through space. The nature of that propagation is expressed by the **Poynting vector**:

$$\mathbf{P} = \mathbf{E} \times \mathbf{H} \tag{3.26}$$

The Poynting vector gives the direction of travel of the wave. From this it can be seen that, if the **E** and **H** vectors are placed on a surface, the direction of propagation will be normal to that surface. The electric field, magnetic field and direction of propagation are all orthogonal, and the magnitudes are related by $E_0 = c_0 \mu_0 H_0$.

3.1.6 Simplifying assumptions

Maxwell's equations are sufficient to understand the field of classical electrodynamics. However, being partial differential equations, they are difficult to evaluate analytically for more than simple cases. It is possible to apply numerical simulation techniques, and there are commercial tools available to do this, but they are limited in scale. They are mostly used

for simulation of relatively simple structures, such as a submarine or a TV antenna, rather than the intricate patterning of an integrated circuit.

Therefore it is necessary to use some simplifying assumptions to make further progress.

3.1.6.1 Near and far fields

When solving Maxwell's equations for wave propagation, it is possible to neglect some or other terms in the integral (a full derivation is beyond the scope of this thesis, but can be found in chapter 2 of Milligan (2005)).

Expressions where r is the distance from the wave source can be split up into terms with $1/r$ dependence which are labelled far-field terms; $1/r^2$ terms which are labelled radiative near-field; and $1/r^3$ terms labelled near-field (Milligan 2005, p. 46).

The fields are continuous, so the boundaries between the three regions are indistinct and different rules of thumb are used to determine them. Common boundaries are (Milligan 2005, p. 56):

$$\frac{r}{L} < 1 \qquad \text{near field} \tag{3.27}$$

$$1 < \frac{r}{L} < \frac{L}{\lambda} \qquad \text{radiative near field} \tag{3.28}$$

$$\frac{r}{L} > \frac{L}{\lambda} \qquad \text{far field} \tag{3.29}$$

where L is the maximum dimension of the aperture and λ is the wavelength. In a traditional dipole, L is some fraction of a wavelength, often taken to be $L = \lambda/4$. Hence a rule of thumb is that the near field is where $r < \lambda/4$.

In the near-field case the average power flow (given by the Poynting vector) is zero, implying a reactive field, while in the far field case real power flow is achieved (Gregson, McCormick, and Parini 2007, p. 33). In the far field, the **E** and **H** fields are perpendicular, obeying equation 3.26. In the near field they are decoupled and can be considered separately, as in the analysis of inductors and capacitors.

In this thesis I mainly consider near-field effects. Working in electromagnetic security, Agrawal, Archambeault, Rao, and Rohatgi (2002a) classify emanations into two categories: *direct emissions*, emissions of secrets at the baseband; and *unintentional emissions*, emissions from secrets modulating other signals.

The size of the receiving antenna controls the optimal attack strategy. A ten metre quarter-wave far-field antenna may be tuned for a 7.5 MHz signal, but owing to its size it cannot be placed close to the source. The inverse-square law means that, at a distance, little power is received compared to a near-field antenna 1 mm from the source.

Direct emissions, which in smartcards are perhaps tens of megahertz, are most suited to the near field, while the higher frequencies of modulated emissions lend themselves to far-field detection. But the baseband emissions of a PCI Express channel at 2.5 GHz are well-suited to a far-field antenna, so this assumption only holds for these limited circumstances.

3.1.6.2 The magnetic field and the Biot-Savart Law

We now focus on the operation of an integrated circuit. For this a full solution or simulation of Maxwell's equations would be too complex.

Considering only the magnetic field allows us to derive a simpler expression for the magnetic field a distance $\mathbf{R}$ from a current carrying element using the Biot-Savart Law:

$$d\mathbf{B} = \frac{\mu}{4\pi} \frac{I\mathbf{R} \times \nabla\mathbf{L}}{|\mathbf{R}|^3} \tag{3.30}$$

A simple example is the field a distance a from an infinitely long straight conductor (Kraus 1991, p. 225):

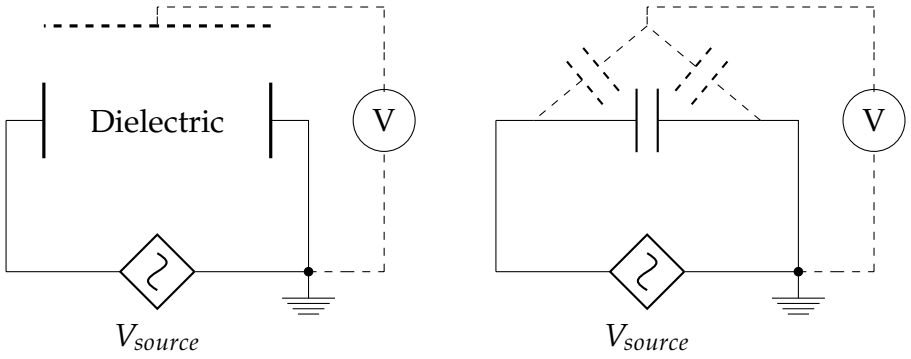
$$|B| = \frac{\mu I}{2\pi a} \tag{3.31}$$

Of interest here is that the spatial and current dimensions are decoupled. That is, at a fixed point from a fixed wire, the only variation of $\mathbf{B}$ is proportional to the time-variation of the current $I(t)$. Of course this assumption will only hold when the frequency rises as far as the sensor stays within the near-field region.

The magnetic field falls off with a $1/r$ relation, suggesting that any sensor should be as close as possible to the current-carrying conductor. The sensor then uses Faraday's Law, or some other effect, to convert the magnetic field to a measurable quantity, such as voltage.

3.1.6.3 The electric field and the displacement current

We can also analyse the electric near field on its own, as a problem of moving charges. Consider the electric field source as forming a capacitor, with



(a) Parasitic capacitor plates (b) Equivalent circuit

Figure 3.1: *Electric field antenna by construction of a parasitic capacitor to an electric field source. The dashed lines indicate the parasitic antenna and receiver added by the eavesdropper.*

a voltage across the plates. We can construct an antenna as another plate, and form another capacitor between the two plates of the field source and the 'parasitic' plate. The capacitor is unlikely to be parallel-plate, and the dielectric is not homogeneous, but it is a capacitor nonetheless. This arrangement is shown in Figure 3.1

The displacement current is given in equation 3.21. Neglecting polarisation, this simplifies to:

$$\mathbf{J_D} = \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \tag{3.32}$$

It is possible to roll up the physical factors into the capacitance C, in which case:

$$I_D = C\frac{dV}{dt} \tag{3.33}$$

where V is the voltage across the capacitor. If there is a time-varying voltage the attacker can receive it through the parasitic capacitors shown in Figure 3.1b. If the capacitor were formed by parallel plates, the capacitance would be given by:

$$C = \frac{\varepsilon_0 \varepsilon_r A}{d} \tag{3.34}$$

where A is the area of the plates and d is the separation distance. This gives us an electric analogue of equation 3.31.

Variations in the E-field may also be caused by dielectric saturation due to the field generated by the IC, analogous to magnetic saturation. This non-linear effect becomes apparent in some dielectrics at $1 \times 10^6\,\mathrm{V\,m^{-1}}$

(Böttcher 1973, p. 289ff), which corresponds to a plate separation of 1 µm at 1 V. This has the considerable advantage of keeping the sensing area small, but it remains to be seen whether it is measurable in silicon.

3.1.7 Uses of the electromagnetic theory

Given these simplifying assumptions, how might we use the tools of electromagnetic theory to progress our understanding and investigation into electromagnetic effects?

First, by considering the electric and magnetic fields separately, we can reduce the complexity by using the relations which relate E- or M- fields into voltages or currents in a sensor that we can measure. This simplifies our understanding of the sensor (into a simple transfer function, rather than a complex three-dimensional sensor requiring numerical analysis). In particular, the circuits and the sensors under test are sufficiently complex and/or unknown that sometimes only qualitative results are possible: they may be modelled to provide quantitative measurements, but such modelling brings a further set of assumptions.

We must be aware of the background, however, to know when the reducing assumptions no longer apply: going beyond the near-field, using a non-static sensor, and so on.

3.2 TEST EQUIPMENT

To perform emissions measurements I needed various laboratory instruments. These arrange the stimuli for the device under test (DUT), record the measurement and then return it to a computer for analysis.

I made both time-domain and frequency-domain measurements. In this thesis I refer to a measurement of some parameter over time using an oscilloscope as a *trace* and a measurement of a parameter over varying frequency using a spectrum analyser or network analyser as a *spectrum*.

3.2.1 Instrument control

Experiments were controlled by a PC running Linux. The Matlab language was used to script experimental procedures, to gather results, and to save and plot them. To control the instruments I used two interfaces: the General-Purpose Interface Bus (GPIB), originally designed by Hewlett Packard, and the VXI11 protocol which is effectively GPIB over Ethernet.

Matlab release R2006a had poor instrument control support under Linux, so I had to implement my own routines. The motivation for sticking with the Linux version was that some of my tests required weeks to run, so it was easier to remotely login and monitor them from a Linux machine than from a Windows one. This saved a great deal of time as I was easily able to restart stalled experiments when away from the laboratory.

First I fitted a National Instruments PCMCIA-GPIB card in a PC Card to PCI adaptor in the PC. Then I compiled the Linux GPIB Project (Hess et al. undated) drivers into the Linux kernel. This provided basic GPIB support for C programs.

I then used the GPIB Module for Perl (Mock et al. 2002) to write some Perl functions that controlled GPIB devices using the Linux GPIB interface. This was added as a backend to a Perl script by Markus Kuhn which controlled various instruments in the laboratory, but was originally designed to use a remote server with the GPIB card.

This stack enabled operations such as 'set frequency' and 'download trace' to instruments from the Linux shell, as well as the sending of lower-level commands. I wrote an interface to provide Matlab scripts such as `spectrumanalyser_downloadtrace` or `spectrumanalyser_trigger`. I also added support for further instruments to Markus Kuhn's Perl script, notably the AFG3252 function generator and the QL335TP power supply.

In addition I wrote an interface to the Linux VXI11 library (Sharples et al. 2008). This library provides GPIB-like calls from C programs, but which are actually sent over TCP using a Remote Procedure Call mechanism. In this case I wrote a C program that uses the Matlab MEX interface to create a shared library that is loaded by Matlab when functions to send or receive VXI11 commands are executed.

Since my experiments involved large quantities of control flow (for example, set the oscilloscope parameters, arm, trigger, read out a trace, repeating thousands of times), I eventually standardised on connecting everything by GPIB as this was more predictable in terms of timing. The X/Y/Z stages, which need only a few bytes to control, were driven by RS232 from the same PC.

All the experiments I undertook were automated. A Matlab script sets up the initial conditions and programs the chip under test with its configuration data. Any instruments providing inputs are configured (voltages or frequencies). Then a measurement cycle is run: the oscilloscope or spectrum analyser is set to trigger, the chip activated, the data captured in the instrument and then read out into Matlab. If averaging is used, the measurement is repeated several times. Then the parameters are changed (the

XYZ stage moved, the voltage or frequency altered) and the measurement is repeated. Each measurement took between a second and a minute. For high-resolution spatial scanning this meant leaving the experiment running for several weeks.

3.2.2 Oscilloscopes

For time-domain measurements I mainly used the Tektronix TDS7254B high-end 'digital phosphor oscilloscope'. It has a 64 MB memory space into which it can record up to four channels at 2.5 GHz at a total of 20 GSamples/s. It contains a PC running Microsoft Windows XP, with a hard drive and full PC connectivity including Ethernet and GPIB. Signal processing functions are provided on the oscilloscope for the calculation of RMS values, jitter, in addition to more specialised functions, such as eye analysis of communications signals.

Being a PC, other software, such as Matlab, can be installed directly on the oscilloscope to read data and control its functions. In these experiments I did not use any, instead I used a separate Linux PC for analysis and control to permit remote login and easier control of the experiments over the network.

The oscilloscope was used with three analog frontend interfaces: the TCA-1MEG high impedance buffer, to which is fitted a conventional oscilloscope probe; the TCA-BNC for 50 Ω inputs over coaxial cables from test equipment and circuits constructed with 50 Ω outputs; and the P7330 differential probe.

The P7330 is a module that plugs into the frontend interface of the oscilloscope and provides a cable with two pins on the end for measuring differential signals. Its specification can be seen in Table 3.1. Many of my sensors produce differential outputs, and the P7330 was used to measure them while excluding large common-mode signals (such as 50 Hz induced mains).

Bandwidth	3.5 GHz
Attenuation	5×
Differential Input Resistance	100 kΩ
Differential Input Range	±2 V
Common Mode Rejection Ratio	60 dB at 1 MHz
Noise	Approx 35 nV/$\sqrt{}$Hz

Table 3.1: *Specification of Tektronix P7330 differential probe*

To drive the spectrum analyser, the TDS7254B was used as a pass-through amplifier. It has an Analog Signal Output which has a bandwidth of 1 GHz into 50 Ω, which was wired with a 30 cm BNC cable and BNC-N adaptor to the spectrum analyser. The gain setting on the differential probe affects the gain of the signal output: at the highest gain of 10 mV div^{-1}, a calibration signal of 0.5 V pk–pk into the differential probe was measured to have unity gain, producing 0.5 V pk–pk on the output port.

I also used other digital oscilloscopes: the LeCroy LC564A (1 GHz) and the Tektronix TDS2024 (200 MHz). These are more limited in their measurement and data collection features.

3.2.2.1 *Trace resampling with time-shift before averaging*

When any digital oscilloscope records a trace, there is uncertainty in the timing. The oscilloscope will record its signal at regular intervals after a trigger signal, but the first sample will not coincide exactly with the trigger edge: there will be some delay. As the system under test is not synchronised with the oscilloscope's internal sampling clock, this delay is different each time a trace is recorded. If multiple traces are averaged together, the averaging will be imperfect since each trace will have a small time-shift. For small differential power/EM analysis measurements, this can be significant.

The TDS7254B measures the time from the triggering event until the first sample (typically a number of picoseconds), and supplies it in the header of a trace download. If the signal being measured has its highest frequency below the Nyquist frequency, it is possible to faithfully reconstruct the waveform by means of interpolation. Such interpolation should be chosen to minimise the additional frequency components (artifacts) that are added to the signal. Once interpolated, new sampling points can then be chosen. This idea was first proposed by Skorobogatov and Kuhn (2005).

If the oscilloscope applies a low-pass filter before sampling the signal, we can trust that the measurement does only contain components below the Nyquist frequency. Then we need to resample in order to shift the phase. In effect this is convolution to form a linear-phase low-pass filter.

In our case, however, processing is happening offline so there is no significant restriction on resources. An alternative approach to resampling is to use a more complex interpolation function, but one which does not require low pass filtering. Figure 3.3a on page 58 shows a sine wave sampled at 7 points per cycle, and various interpolation strategies applied to upsample it 40 times. Figure 3.3b on page 58 shows the amplitude of their

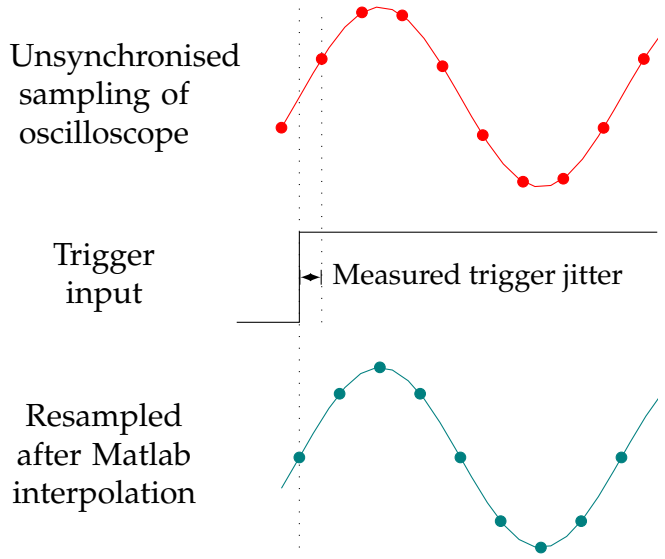


Figure 3.2: *Resampling oscilloscope traces to remove timing jitter after trigger signal*

spectra. The mathematically more complex interpolation schemes show a much lower level of additional harmonics. Ideally a sinc interpolation function would be used but Matlab does not provide a convenient interpolator: its spline interpolator works sufficiently well in practice.

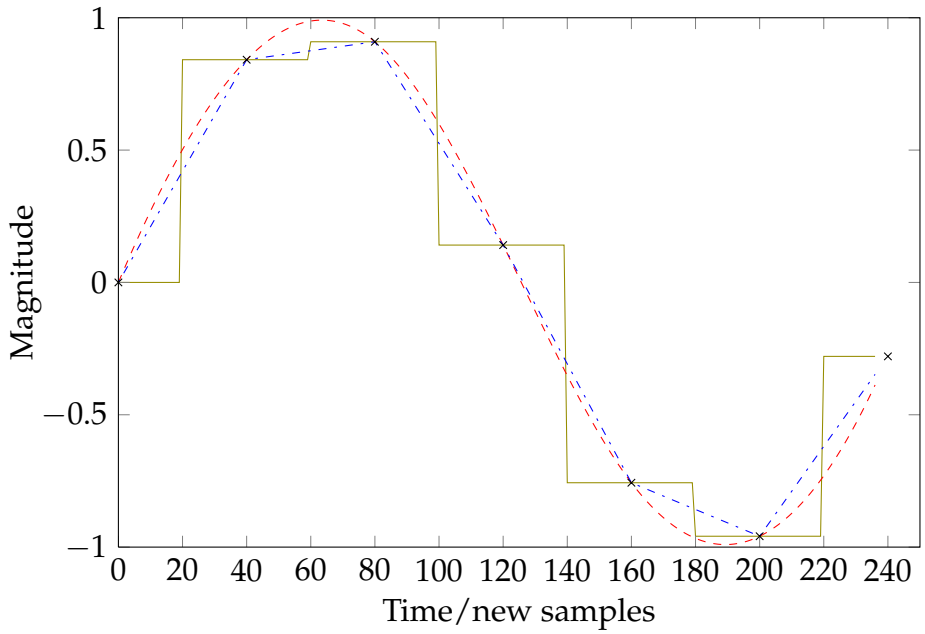
For the oscilloscope data, the script `resampletimetrace.m` will create a new time series starting at the trigger point, and interpolate the recorded trace onto this new sampling basis using a spline function as that produces the fewest additional spectral components. Thus all modified traces use the same sampling basis.

3.2.3 Spectrum analyser

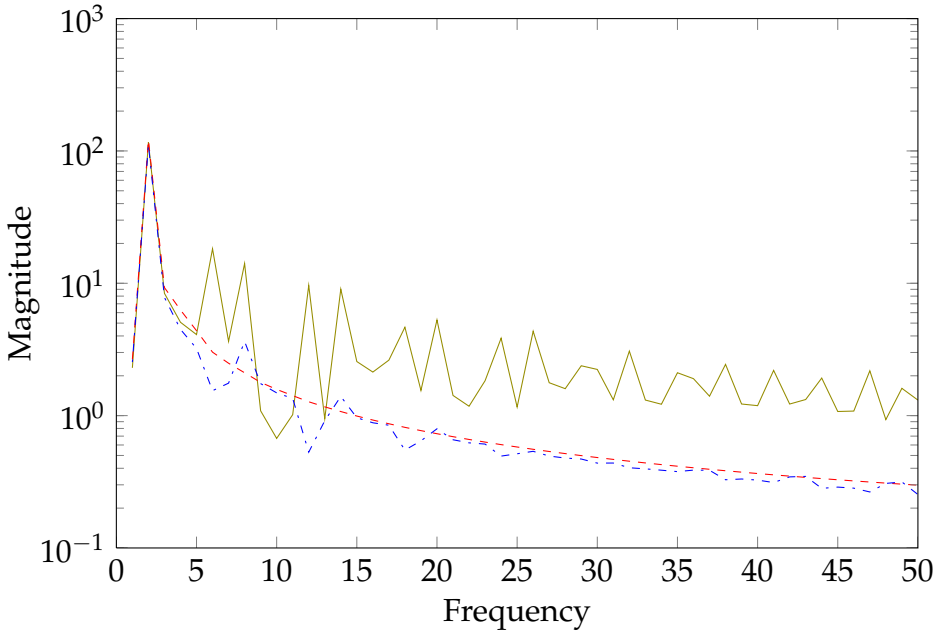
For identifing the spectral components of a measured signal, I used an Anritsu MS2601B spectrum analyser. This is a microprocessor-controlled device with a digital capture of measurements: key details from the specification of this instrument are outlined in Table 3.2 on page 59.

Older spectrum analysers[2], such as the model I used, are built from a

[2]Modern spectrum analysers use analogue hardware only for downconversion, then feed the digitised output into a software radio and use digital signal processing to per-



(a) Time domain



(b) Frequency domain

Figure 3.3: *A sampled sine wave (black crosses) has been upsampled by 40 times by interpolation: nearest-neighbour (solid ——), linear (dot-dash - · - ·), spline (dashed - - -). The frequency spectrum indicates that the rougher interpolation methods introduce unwanted extra harmonics.*

Frequency range	9 kHz to 2.2 GHz
Input impedance	50 Ω
Amplitude range	−120 to 20 dBm
Maximum displayed average noise	−120 dBm
Maximum dynamic range	−75 dB
Resolution bandwidth	30 Hz to 1 MHz
Video bandwidth	1 Hz to 100 kHz
Sweep time	50 ms to 1000 s
Display resolution	20 Hz

Table 3.2: *Anritsu MS2601B spectrum analyser specification.*

tuned radio receiver which is swept through a range of frequencies. At each discrete frequency the amplitude of the received signal is measured and a plot of frequency against amplitude produced. The MS2601B displays its amplitude measurements in several units. In my measurements I used dBm, or decibels relative to one milliwatt. This is a measure of power, given by:

$$p = 10 \log_{10} \frac{P}{P_0} \tag{3.35}$$

where P is the linear power (in watts), P_0 is the reference power and p the logarithmic power (in decibels). A reference power of $P_0 = 1\,\text{mW}$ gives p in dBm (in other words, 0 dBm = 1 mW).

The spectrum analyser provides a number of settings to adjust the measurement process. The instrument has a built-in attenuator. This was set to 10 dB, i.e. with 10× attenuation (the spectrum analyser has a lower 0 dB attenuation function but will only select it in some modes). Since a spectrum analyser works using a frequency sweep, it is impossible to sample all frequencies at the same instant. The *sweep time* setting allows adjustment of the time taken to sample all the frequencies in the measurement range.

The *resolution bandwidth* (RBW) is the bandwidth of the Intermediate Frequency amplifier. Thus each point n, or *bin*, on the recorded spectrum $X(f_n)$ is the integral of the spectral components in a band between $f_n \pm \frac{1}{2}\text{RBW}$. A larger resolution bandwidth means each frequency bin receives more energy, settles faster and so provides a faster spectrum capture, but at the expense of not discriminating frequencies close together. The RBW was adjusted for each frequency range to give sufficient resol-

ution for the measurement being undertaken, mostly using the automatic RBW function. For sweeps which involved a wide range of measurements at small frequency intervals (for example, the Lochside distributed clock generator experiments in Section 4.7.2 on page 118) the RBW was adjusted for each measurement.

After amplitude demodulation, the *video bandwidth* (VBW) filter is applied to the recorded trace before it is presented on the screen (or recorded in memory and transmitted to the PC). This allows high frequency variations between nearby frequency bins to be filtered out, hiding some elements of noise in the captured signal. The effect depends on the ratio VBW:RBW; a small ratio implies more filtering. For my application these local variations are interesting and I did not wish to filter them out. I either disabled the VBW filter completely or used the spectrum analyser's default settings, which has a large ratio and thus little filtering.

Being a digital device, the trace is captured to local RAM and can be read out by the PC over GPIB. Additional processing can take place here, with the instrument averaging several sweeps. In my experiments I recorded and downloaded individual sweeps and performed processing myself in Matlab.

When I draw spectral plots, which are typically averaged over 20 spectra, I display the mean spectrum and spectra showing a standard deviation each side ($\pm\sigma$). In many cases these are very close together, indicating the repeatability of spectral measurements.

3.2.3.1 *Differential spectral analysis*

Analogous to differential power analysis, I derived *Differential Spectral Analysis* (DSA). The premise is, just as DPA works in the time domain, DSA works by taking the differences between spectra in the frequency domain. As with my minimal implementation of DPA (taking the difference of SPA traces), to implement DSA I simply subtracted two traces and looked at any variations in peaks, particularly in frequency differences (less likely to be due to noise than amplitude differences). In principle the same statistical techniques to extract data from noise in DPA would apply to DSA.

Independently Gebotys, Ho, and Tiu (2005) describes a very similar technique but provides a more thorough evaluation. Their approach differs in that they record the time domain traces and then take their Fast Fourier Transform to compute the spectrum which they then use to perform DPA. This is more accurate than using a spectrum analyser, as it ensures the data capture takes place during the cryptographic operation, but

potentially exposes the results to artifacts of the FFT (such as windowing).

The DSA approach removes much of the need for shielding, or an expensive test chamber. The location of the experiments was about 4 km as the crow flies from Madingley Transmitter, which broadcasts UHF analogue TV on 575.25 MHz at up to 5 kW (the 'Five' channel) and various FM radio stations on VHF, most notably BBC Radio Cambridgeshire on 96.0 MHz at 1 kW (Brown 2010a, Brown 2010b, Queens' College, Cambridge 2000). In addition the controlling PC leaks some frequencies, such as the 133 MHz bus clock, and other frequencies are present due to the test location in a building full of computers and separated by two plasterboard walls from a server room. Some of these signals can be seen on the recorded spectra but, being constant, these are mostly removed by DSA.

3.2.4 Vector Network Analyser

A *network analyser* is a device which measures the response of a two- or four-terminal passive or active network according to frequency. In a four-terminal network, the network analyser injects a signal between two terminals of the network and measures the signal received between the other two terminals, i.e. the component of the input that has been transmitted from the input pair to the output pair. In the two-terminal case, the network analyser injects a signal into the network and measures the amount received back, i.e. the reflected component.

A *Scalar Network Analyser* (SNA) measures the magnitude of the response of the network at each of a range of frequencies. It can be constructed with a spectrum analyser and a *tracking generator*: a linked signal generator which injects the frequency the spectrum analyser is currently measuring.

A *Vector Network Analyser* (VNA) measures the phase of the network as well as the gain. The reflected and transmitted coefficients in both gain and phase given the network scattering parameters, or *S-parameters*, which describe the behaviour of the network. These may be displayed graphically in the complex plane using the *Smith Chart*.

Typically, a broadband VNA is constructed using a Voltage Standing-Wave Ratio (VSWR) bridge, similar to a Wheatstone bridge where the device under test (DUT) takes up one arm. When an AC voltage is applied to the bridge, some of that signal appears across the DUT. If the DUT does not perfectly terminate the bridge, signals proportional to the incident and

Frequency range	9 kHz to 4 GHz
Output power	3.2 µW to 1 mW
Output impedance	50 Ω
Dynamic range	95 to 120 dB
Displayed average noise	−100 to −75 dBm

Table 3.3: *Rohde & Schwartz ZVRE vector network analyser specification*

reflected voltages appear across resistors in the bridge. Their amplitude and phase can be measured to determine the impedance of the DUT at the injected frequency.

I used a Rohde and Schwartz ZVRE vector network analyser for some measurements. Its specifications are outlined in Table 3.3.

In many cases, however, a VNA did not give a useful result. When I wanted to measure the emissions from an on-chip antenna the VNA would not help, since any attempt to drive the on-chip antenna from the VNA would involve routing the signal down a wire, into the pin of the chip, along the bond wire, through the I/O pad and around the chip metallisation before reaching the antenna. When not being measured, the signal would not otherwise take this path, and emissions from the wiring were found to dwarf any small signal emitted by the antenna.

Therefore I developed my own SNA. An on-chip oscillator was used to generate a signal at a known frequency and this was used to drive the antenna. A spectrum analyser was used to measure the emissions from the antenna in a band centred on the given frequency. This measured the frequency response of the oscillator and antenna combined. Since these are on a similar scale, the antenna may be evaluated by comparing with just the oscillator running. In the VNA case this would be difficult, since the antenna difference is lost in the noise of the other VNA emissions.

3.2.5 Other equipment

I also used a Tektronix AFG3252 arbitrary function generator to generate different waveforms up to 240 MHz, and in a few cases an Agilent E4421B signal generator for sinusoids up to 3 GHz.

For power supplies, variously the Farnell TOPS 3D, the Digimess HY3003-2, the TTi PL330DP and the TTi QL355TP were used when available (they were shared with other users of the laboratory). The last was used under GPIB control, the others were set manually. Up to three separate power supplies were used in a test. One powered the boards under

test, where both Springbank and Lochside have multiple internal supplies and onboard voltage regulators (for the Lochside board I designed a software-adjustable core voltage regulator). For active sensors, another provided one or two voltages to power the sensor amplifier. Finally a third was used as a precision stimulus for the DCG control voltage used in Section 4.7.2 on page 118). The separation of supplies was to minimise signal coupling via the power supply instead of via the sensor.

For positioning, Newport Instruments X/Y/Z stages were used as described in Section 4.7.4 on page 127.

3.3 TEST BOARDS

I used a selection of processor boards for testing purposes. One was an off-the-shelf 32-bit ARM microcontroller (LH77790B), which I used for preliminary experimentation simply because it was to hand.

Another was a chip (Springbank) containing a number of 8-bit microcontrollers that had been designed for a previous research project, and had been subjected to a previous security analysis. The reason for this choice was that all the design files, layout and documentation were available, and so the behaviour of the device was well understood. This is very different from using a commercial device where, even if depackaged, the best that can be expected is a photo of the die and details on the application-oriented datasheet. In such a case even determining the technology of fabrication is nontrivial, and a large amount of guesswork is required.

I was also involved in fabrication of a third chip (Lochside) that was intended for another research project, and was able to design and integrate some test structures on one edge of the die. Due to the very tight timescales I had no time to optimise these structures. They were originally intended to be used to measure the effects of different semiconductor layout techniques, but it turned out that these were too small to be measurable with my equipment. Thus the clock generation circuits became of more interest than the layout structures. I also designed much of the support hardware on the Lochside evaluation board.

Finally for random number experiments I also used a commercial 8-bit microcontroller on its evaluation board intended for security-sensitive applications, that has been and continues to be used in banking terminals. Again, only the user-facing datasheet was available. Due to a nondisclosure agreement, the owner of the board preferred that I not publish the model of the device.

I also constructed a number of supplementary boards, and adapted

some commercial hardware such as smartcard readers. These will be described where relevant in the text. The majority of sensing hardware was of my own design and construction.

3.3.1 LH77790B test board

An ARM AEB1 evaluation board was used to experiment with some sensors. This contains a Sharp LH77790B ARM7DI processor clocked at 25 MHz, as well as 128 kB of off-chip RAM and 256 kB flash (Figure 3.4).

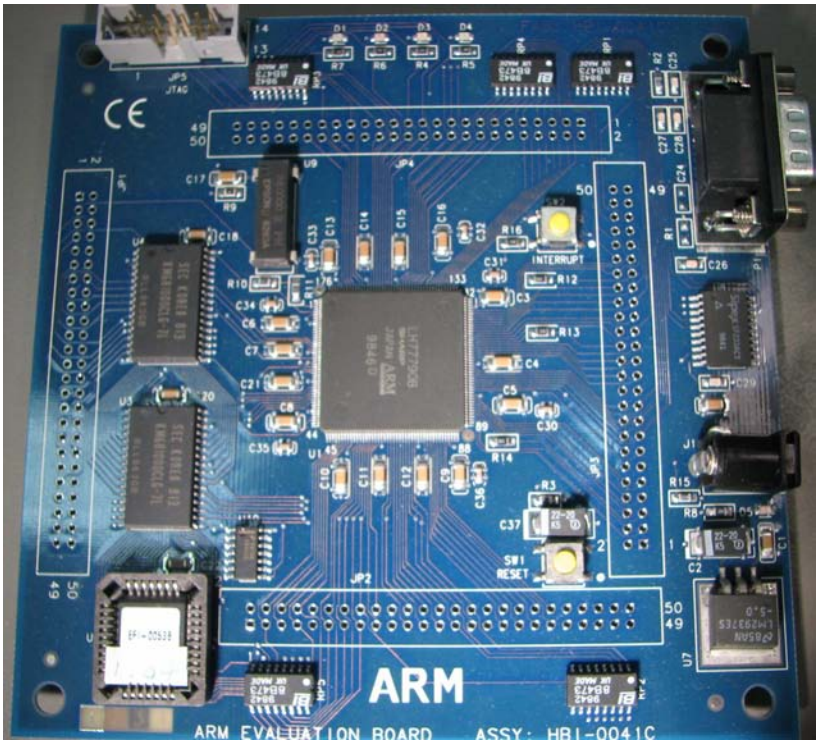


Figure 3.4: *ARM AEB1 test board with Sharp LH77790B ARM-based processor*

The CPU ran the following loop:

```
; r0 points to an output port
; r3 contains a preloaded value 3
loop
        EOR r1,r1,#1          ; toggle trigger bit
        STRB r1,[r0]          ; write bit to an output port
        SUB r2,r3,#2          ; r2=r3-2
        B loop
```

The ROM monitor and conventional interrupts (IRQs) were disabled by

the initialisation code. A FIQ (fast interrupt) handler was registered so that pressing a button would generate an interrupt and change the value of R3 from 3 to 1 or vice versa. With R3=3, the subtract would give the result 1 (Hamming weight 1), while, with R3=1 the result would be -1 (Hamming weight 31). The loop code and register values remained unchanged except for R2 and R3. All the while a square wave is generated on an I/O port to allow triggering an oscilloscope.

A second test involved running three tight loops with different operating frequencies and looking for correlated periodicities in the power/EM trace:

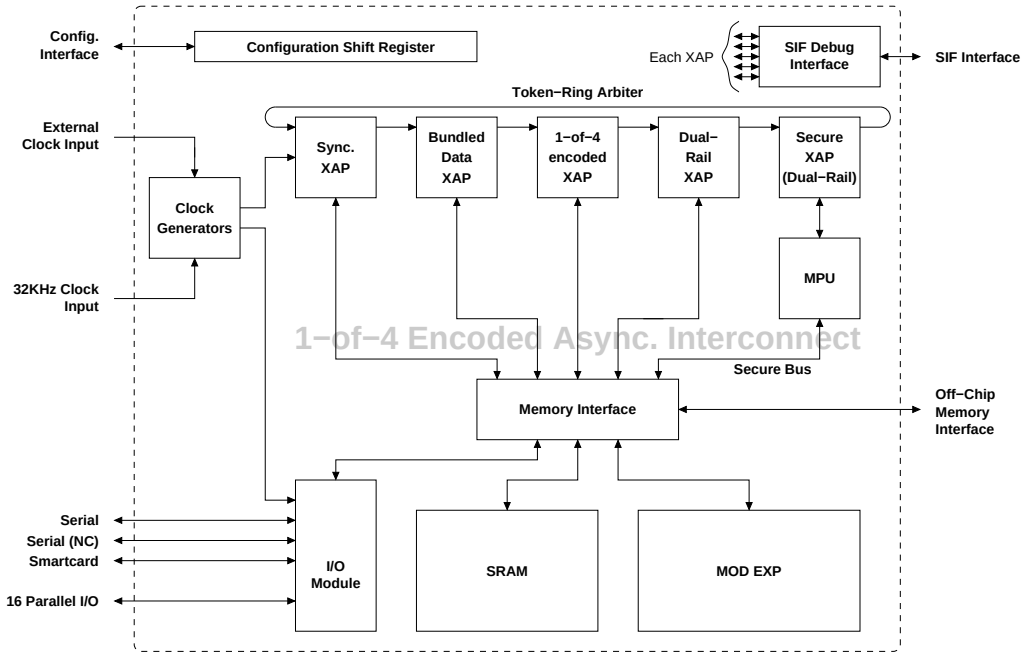
```
loop1
        B loop1

loop2
        AND r0,r0,r0
        B loop2

loop3
        TST r3,#2     ; creates data dependent timing
        BEQ loop3
        MOV r0,r0
        B   loop3
```

The results for different sensors may be found in Chapter 4.

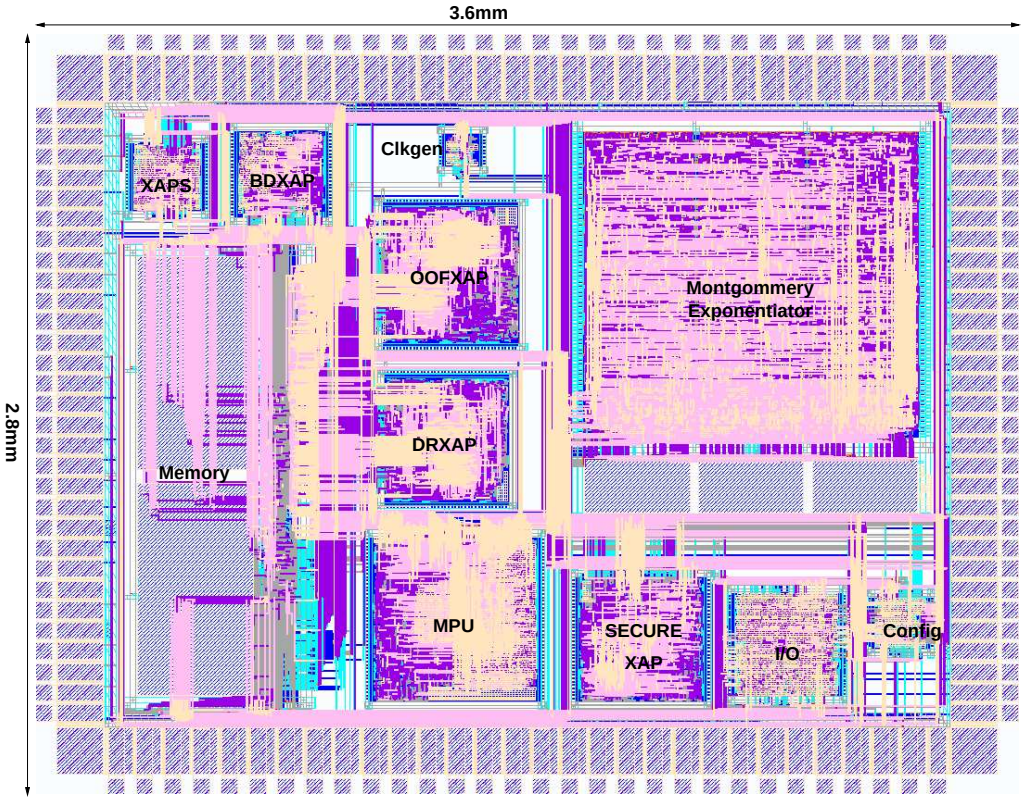
3.3.2 *Springbank test chip*

For a test subject I used the Springbank chip (Moore, Anderson, Mullins, Taylor, and Fournier 2003), fabricated as part of the G3Card project (G3Card Consortium 2003). This comprises five 16-bit XAP1 processors in a variety of design styles: synchronous, bundled data, 1-of-4 encoded, dual-rail and 'secure' self-timed processors. The Secure XAP also uses memory protection and bus encryption. 4096×18 bits of SRAM are provided, as well as general purpose I/O, two asynchronous UARTs, a smartcard interface, and a (non-functional) true random number generator. A Montgomery modular exponentiator is built in self-timed dual-rail logic. Clocks and a SIF debugging interface are also provided. A block diagram and die plot may be seen in Figure 3.5 on the following page.

The SIF bus allows programs to be run on a chosen XAP processor, making use of the other peripherals. Therefore we can compare emissions from one XAP design style to another.

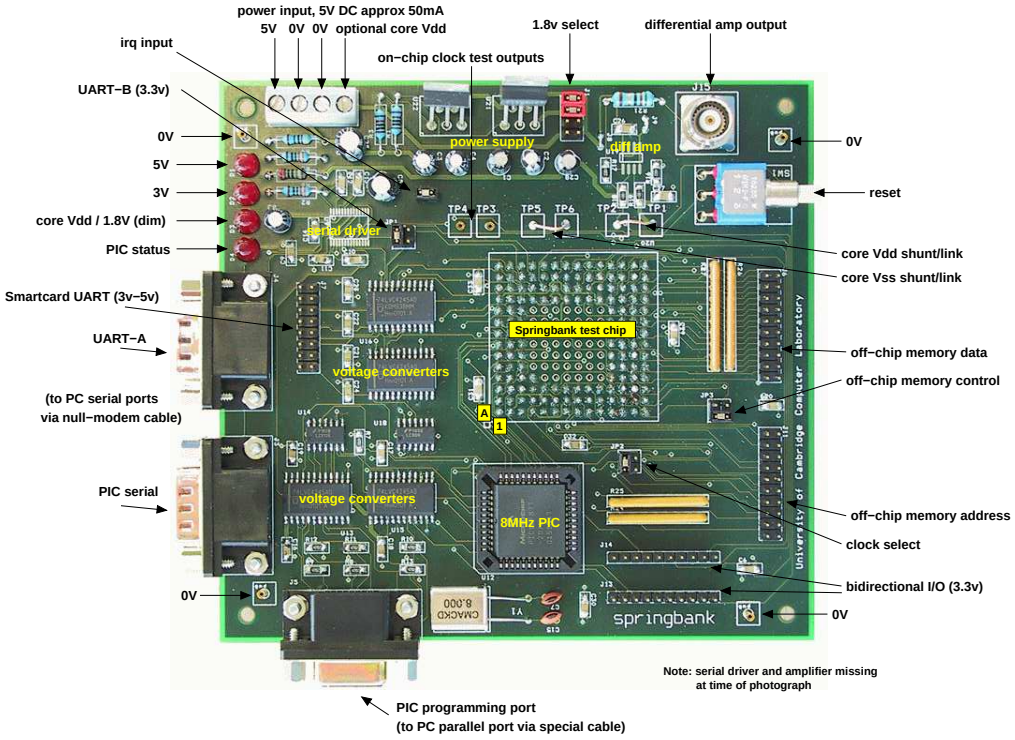


(a) Block diagram

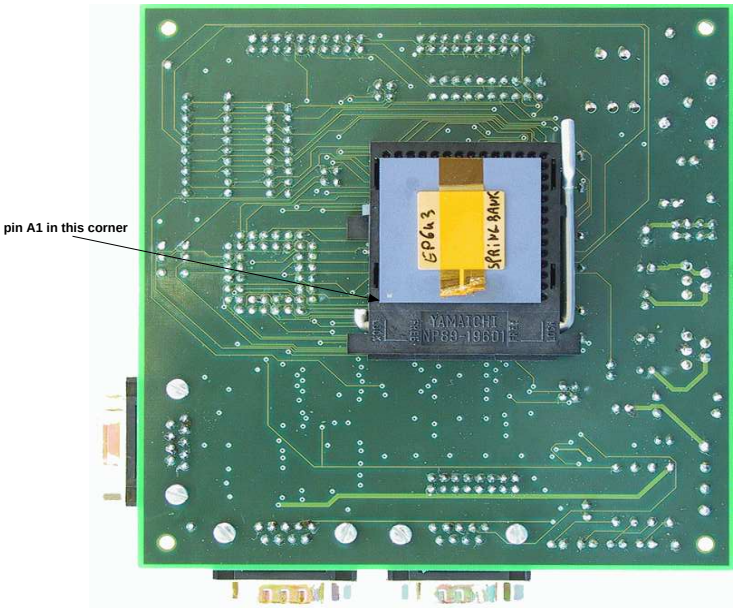


(b) Layout

Figure 3.5: *Organisation of the Springbank chip*



(a) Upper side



(b) Reverse side

Figure 3.6: *The Springbank board*

The Springbank chip is fabricated in a 0.25 µm UMC 6-metal-layer process. The Springbank board (Figure 3.6 on the previous page) has the chip mounted on the rear side, so that it can be easily probed by sensors which need to be close to the die surface, such as a hard drive head in an X/Y/Z micropositioning system. A PIC microcontroller on the board loads code and data into the Springbank chip, and is controlled over an RS232 serial cable with a Java program on a PC.

A previous security analysis (Fournier, Moore, Li, Mullins, and Taylor 2003) found the Springbank to have power and E-M leakage due to mismatches in the commercial memory macrocell. This makes the Springbank interesting as a target since it is both designed as a security circuit but known to be vulnerable.

3.3.2.1 Electromagnetic simulation

Huiyun Li and I developed a simulator to calculate electromagnetic emissions of a circuit, based on a silicon layout (Li, Markettos, and Moore 2005). This enables the designer to evaluate some of the likely emissions properties of their proposed design before it is fabricated, and enables them to design for optimal emissions. The simulation breaks the design down into blocks of interest, then computes their instantaneous power consumption over the course of an operation of interest (such as a DES round) by measuring the current and voltage in a circuit loop. This, applied to a model of a sensor and, optionally, its modulation/demodulation function, provides traces which may be used in a simulated power analysis attack.

In conjunction with these simulations I performed some EMA measurements on Springbank to confirm the validity of the simulation. These results may be found in Section 4.6 on page 111.

3.3.3 Lochside test chip

To enable calibration of electromagnetic sensors, I designed some test circuits. The aim was to design a device containing different circuit structures whose EM emissions could then be measured and compared. These were fabricated as part of the Lochside chip (Mullins, West, and Moore 2004), a 5×5 mm ASIC in 0.18 µm UMC technology with six metal layers. The main part of the Lochside chip comprises a network on chip with 16 nodes, taking up 4.4×4.4 mm. Embedded in this is a distributed clock generator based on asynchronous circuit techniques (Fairbanks and Moore 2005). My security test block is along one edge, occupying 3.2×0.1 mm. Also

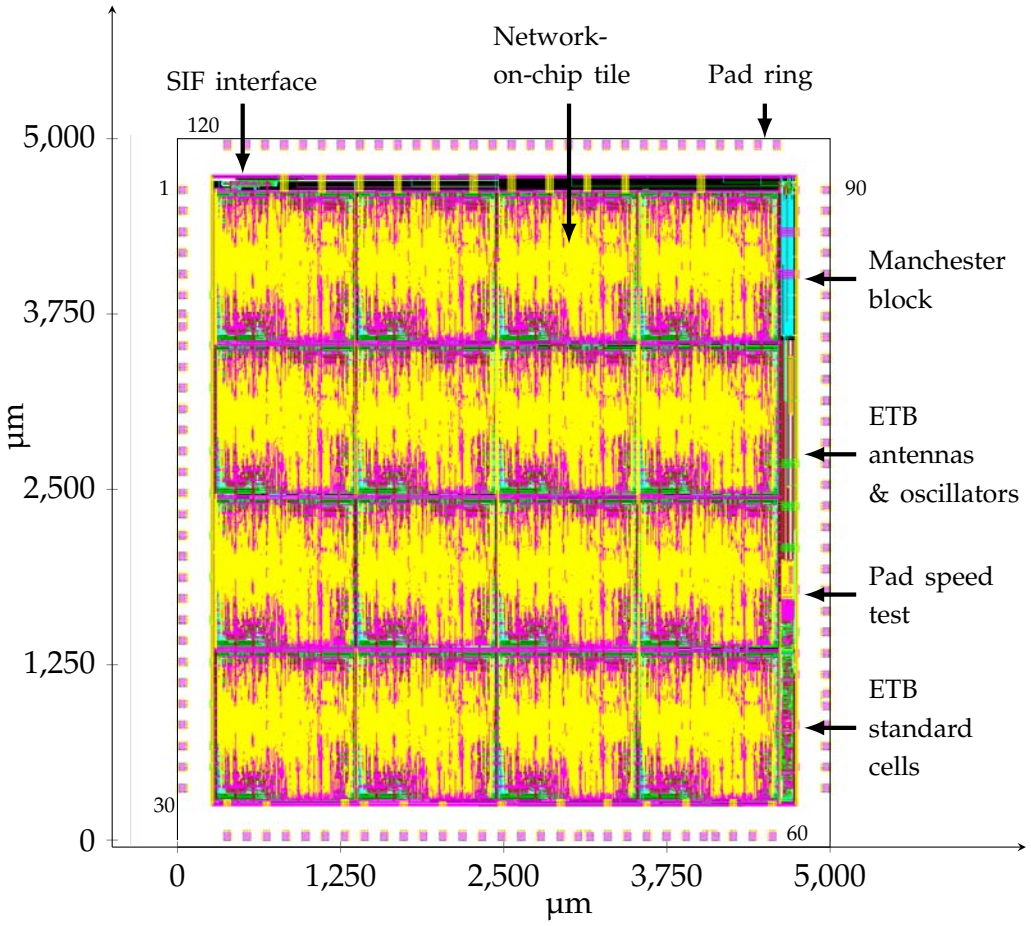


Figure 3.7: *Lochside chip layout*

present is a circuit to measure the speed of the I/O pads and an asynchronous ternary logic block designed at the University of Manchester, which I do not use. A plot of the layout may be seen in Figure 3.7 and the pinout in Figure 3.8 on the following page. Full design details of the security block may be found in the datasheet in Appendix A.

The Security test block is comprised of the structures shown in Figure 3.9 on page 71. A series of different clock generators may be connected to a variety of wire lengths and loops via a clock divider.

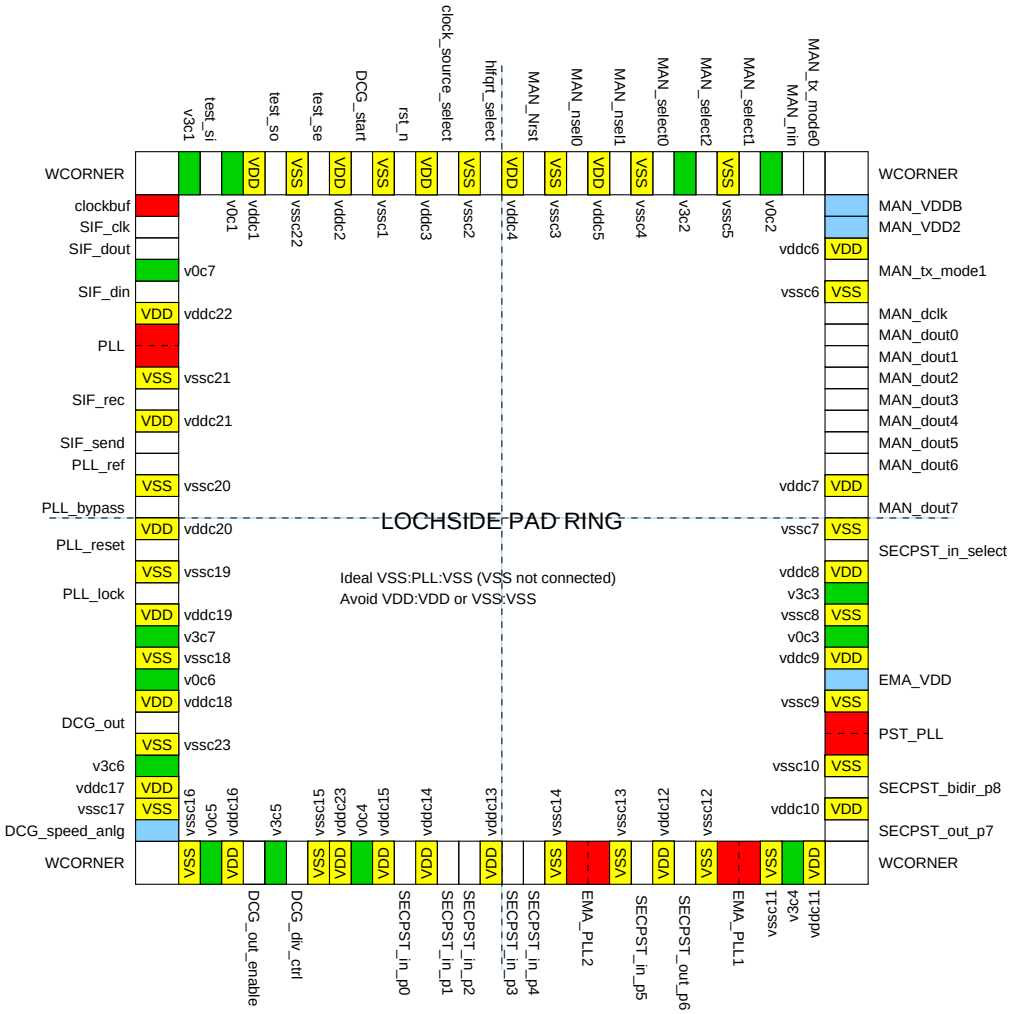


Figure 3.8: *Lochside pad ring (orientation same as Figure 3.7 on the preceding page)*

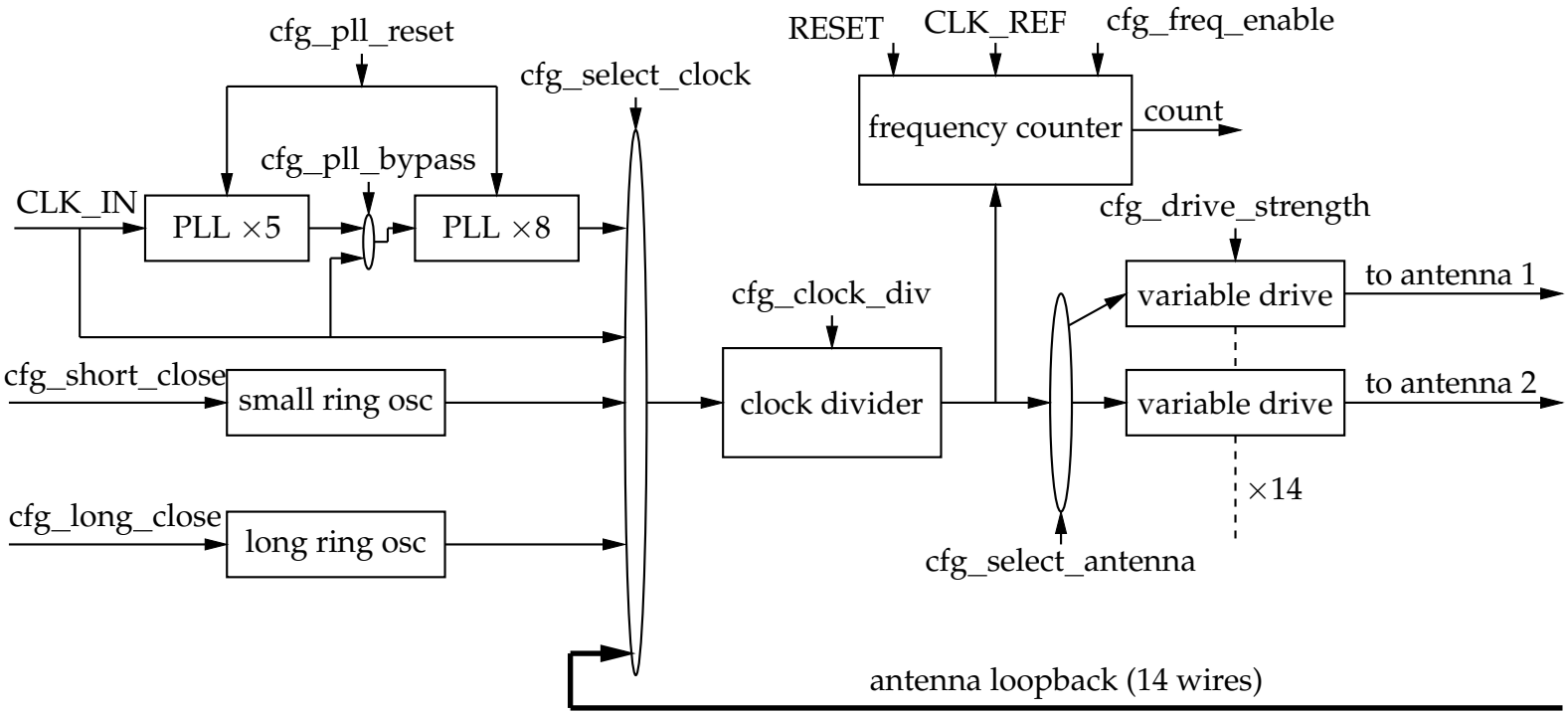


Figure 3.9: *Lochside Emissions Test Block organisation*

Antennas are driven with a variable strength driver. This allows the drive effort applied to the antenna wire to be powers of two between 1 and 128 times the drive strength of a single 'unity' transistor (width 2.56 µm).

The antenna designs are described in Table 3.4 on the facing page: further details of the layout may be found Appendix A.

The chip was packaged in a 120-pin Pin Grid Array package. A test PCB (Figure 3.10 on page 75) was designed and fabricated, having a ZIF socket on the reverse side to aid probing the chip. The PCB contains a Texas Instruments TUSB3210 microcontroller which provides a USB interface. The firmware on the microcontroller enables the control of all the input pins to the Lochside chip from commands sent over USB, as well as other routines. The security test block has a separate power input, `VDD_EMA`, to enable power analysis over just the security circuits. The PCB provides a digitally-adjustable power supply, 0 to 2.0 V, to drive `VDD_EMA`. The PCB also provides fixed voltage regulators (5 V, 3.3 V, 1.8 V) for the board and chip (the Lochside I/O runs at 3.3 V while the core runs at a nominal 1.8 V). A crystal and PLL-based clock generator are also supported on the PCB. These were removed from the board for EMA tests since they produced too many emissions of their own.

3.3.3.1 Test results

The EMA test block on the Lochside chip was found to operate as intended, with no bugs in the logic operation. Measurements of the functions are described as follows. Where a configuration value is given, this corresponds to the register layout described in Appendix A.

With a core power supply of 1.8 V, the long ring oscillator and clock divider gave the outputs on `CLK_OUT` seen in table Table 3.5 on page 74. The long ring oscillator scales in frequency linearly with supply voltage (Figure 3.11 on page 76). In other experiments a core supply of 2 V was used, resulting in a `CLK_OUT` of about 6.2 MHz. Both are temperature dependent.

When the antennas were formed into closed-loop ring oscillators, results as measured from `CLK_OUT` may be seen in Table 3.6 on page 74. In particular, antennas 2, 3, 14 and 15 did not oscillate, since they do not form closed loops. The effect of increased drive strength was noticeable: antennas 0, 1, 4 and 5 oscillated at about 185 MHz with a drive strength of 128, but at about 115 MHz with a drive strength of 1. Otherwise the exact configuration of metal layers seemed to have a small effect, but much less than active inverters being placed in the loop. Measurements of the electromagnetic behaviour may be found in Chapter 4.

Antenna	Structure
0	1.55 mm metal 6 loop separated by 5 µm, terminals at edge
1	1.55 mm metal 6 running over return path of 1.55 mm metal 5 with connecting via at far end
2	1.55 mm parallel metal 6 wires, separation 5 µm, one driven one grounded and near end
3	1.55 mm metal 6 running over 1.55 mm metal 5 grounded at near end
4	500 µm metal 6 shield over 1.5 mm metal 5 loop, 5 µm wide.
5	1.05 mm metal 6 shield over 1.00 mm metal 5 loop, 5 µm wide
6	1 mm metal 6 loop separated by 5 µm
7	Metal 4 and 6 grounded sandwich around 1 mm metal 5 loop, 5 µm wide
8	1.05 mm metal 5 shield under 1.00 mm metal 6 loop, 5 µm wide
9	1.56 mm × 0.9 mm metal 6 loop around outside of cell
10	100 µm metal 6 loop, 5 µm wide
11	100 µm long loop of 137 inverters, between metal 4 and 6 shielding, 5 µm wide
12	10 µm long loop of 15 inverters, half under metal 6 shielding, 5 µm wide
13	10 µm metal 6 loop, 5 µm wide
14	Output from long ring oscillator

Table 3.4: *Lochside antenna structures: a schematic of these layouts is shown on page 215*

The PLLs are capable of producing an output of up to 850 MHz, as seen in Table 3.7 on page 76. If `pll_bypass`=0, a $\times 40$ multiplier of the PLL input frequency applies. With `pll_bypass`=1, the $\times 5$ PLL is switched out, leaving only the $\times 8$ in operation.

As the fundamental increases the pads are filtering higher harmonics, but the full frequency is used inside. Increasing the input frequency further causes a broader output spectrum (even after division), suggesting the internal frequency is failing some timing constraint, and disappears after about `CLK_IN` $>$ 22.2 MHz i.e. `CLK_OUT` $>$ 888 MHz. The pads are capable of driving at up to about 450 MHz at reduced drive, but not at the

clock_div	Frequency/kHz
0	5680
1	2842
2	1421
3	710.6
4	355.3
5	177
6	88.8
7	44.4

Table 3.5: *Lochside long ring oscillator after clock divider, at* V_{core} =1.8 V*, configuration* `70B63x`

Antenna	Strength	Config.	Frequency /MHz	Amplitude pk-pk /V	DC level/V
0	128	702240	190	0.66	1.44
0	1	002240	113-7	0.9	2.08
1	128	712250	180	0.55	1.48
1	1	012250	113-7	1	2.02
2	128	722260	0	0	0
3	128	732270	0	0	3.3
4	1	742280	113	1	2.07
4	128	752290	182-7	0.59	1.44
5	128	052290	183-7	0.6	1.39
5	1	7622A0	118	0.89	2.1
6	1	0622A0	118	0.88	1.9
7	1	0722B0	108	0.96	2.05
8	1	0822C0	112	1.03	1.98
9	1	0922D0	114	1	2.08
10	1	0A22E0	122	0.55	1.91
11	1	0B22F0	58	1.56	1.77
12	1	0C2300	122	0.48	1.97
13	1	0D2310	131	0.68	2.03
14	1	0E2320	0	0	0
15	1	0F2330	0	0	0

Table 3.6: *Lochside antennas used as loopback oscillators as measured at* `CLK_OUT`*, at* V_{core} =1.8 V

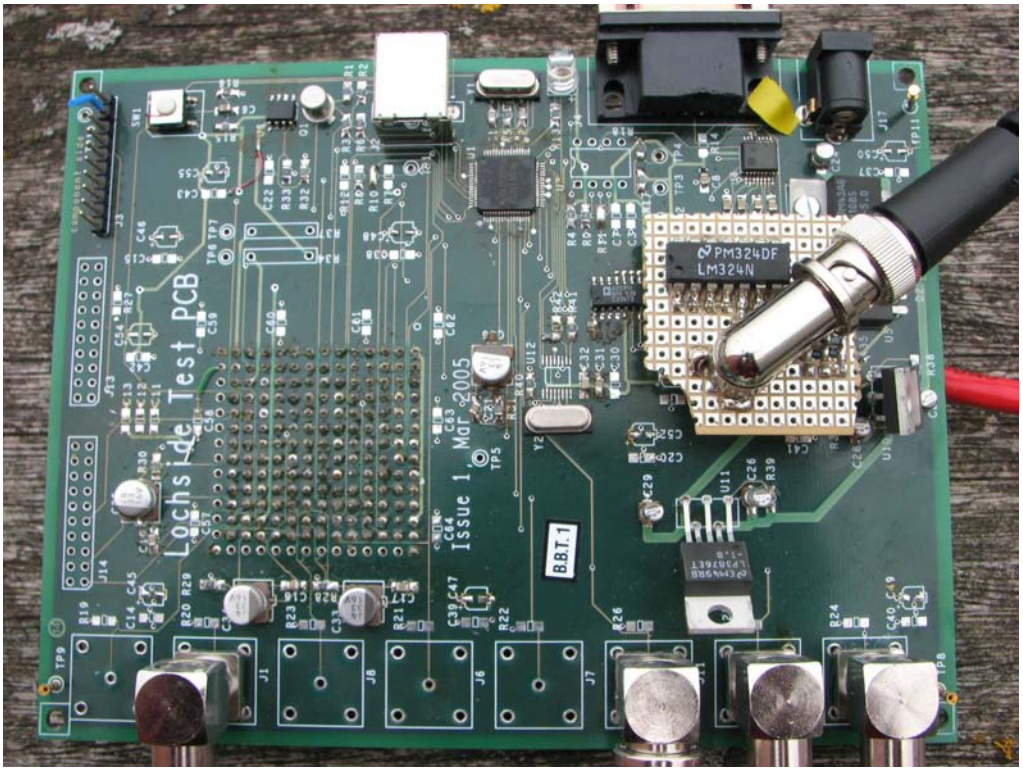


Figure 3.10: *Lochside test PCB. The Lochside chip is mounted on the rear side of the board (Figure 4.5 on page 83). The daughterboard provides the coupler for power supply injection (used in Chapter 5).*

900 MHz undivided output of the PLL.

A Distributed Clock Generator based on asynchronous C-element rings was also present to clock the network-on-chip. This is described further in Section 4.7.2 on page 118. The emissions performance of the Lochside chip is described in Chapter 4.

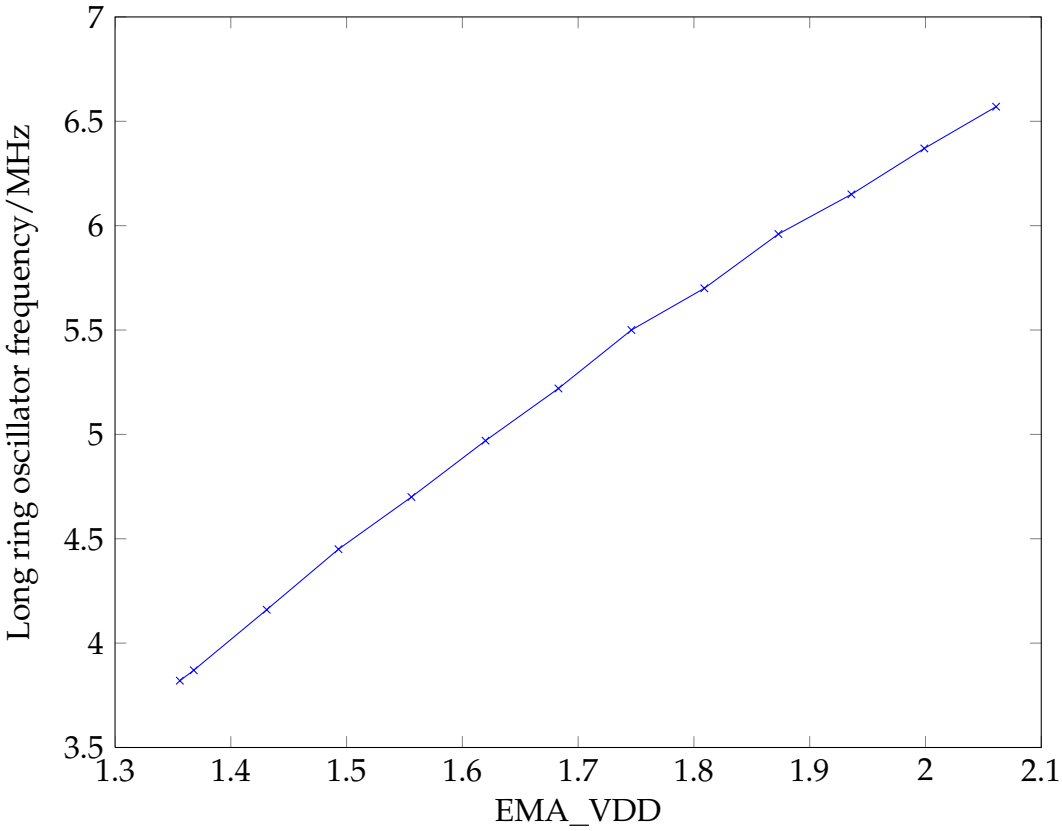


Figure 3.11: *Lochside long ring oscillator variability with supply voltage*

clock_ div	pll_ bypass	Config	PLL input /MHz	Primary har- monic /MHz	Output shape
128	0	00B417	21	6.3	Stepped, 3.7Vpk-pk
64	0	00B416	21	12.4	Stepped, 3.0Vpp
32	0	00B415	21	24.8	Stepped, 2.7Vpp
16	0	00B414	21	49.8	Stepped, 2.2Vpp
8	0	00B413	21	99.6	Triangle, 2.1Vpp
4	0	00B412	21	199.2	Sine 1.0Vpp
2	0	00B411	21	424.8	Sine 0.5Vpp
1	0	00B410	21	0	No output
128	1	00BC17	27.9	1.75	Square, pk-pk 3.4V

Table 3.7: *Lochside PLL operation, at* V_{core} =2.0 V

CHAPTER 4
SENSORS FOR EMA

4.1 INTRODUCTION

Power analysis is performed very simply: adding a resistor into the circuit loop in which current flows between the power supply and the chip. In theory, and often in practice, the position of this resistor has relatively little effect on the measurement being taken, other than practical considerations (such as the reference level of the measuring oscilloscope or other device). With careful choice of resistor and construction, the measurement is mostly frequency-independent for most frequencies of interest. The recording is, for the most part, a function of the actual current consumption of the chip, time, the power supply voltage, the effectively-constant resistor value, and the frequency cutoff of the measurement equipment. The last three can often be neglected, resulting in an approximately two-dimensional parameter space (time, current).

Electromagnetic measurements bring in several more dimensions that must be considered. Firstly, electromagnetic radiation occupies space. That means that an EM measurement depends on the three spatial dimensions. Any measurement is an integral of the field over some surface or volume: it is not possible to measure the field at a single point. Secondly, any EM sensor has a frequency response. EM sensors with a flat, broadband, response are difficult to construct. Thus the measurement contains all the variables of the power measurement but with four additional dimensions which cannot be simplified away. Thus the EM parameter space is approximately six-dimensional (spatial X, Y, Z, frequency, time, current).

Furthermore, even if some of these parameters are fixed, construction of a sensor to satisfy a subset (for example, fixed XYZ, flat frequency 0 to 1 GHz) is non-trivial. Therefore I elected to turn the question around, taking some existing EM sensors and evaluating their properties to decide which might be best-suited for electromagnetic measurements from integrated circuits.

In addition, the search space is large. When no signal can be discerned from a sensor, it could be for many reasons. It could be that the sensor is improperly designed, constructed or is malfunctioning. The design of the experiment and the equipment chosen may be non-optimal. The positioning of the sensor might be imperfect. The sensor might have insufficient gain, or insufficient gain at the frequency of interest. The interesting signal

may be overlayed by uninteresting components. The signal processing of the sensor output might be insufficient to distinguish a signal from noise. The measurement chain may reduce the signal to noise ratio (SNR) rather than increase it. The device under test may not be radiating in the first place.

This means that experimentally exploring this search space is difficult. Furthermore, while some control experiments are possible (such as measuring a test E-M source with known characteristics) they may not be representative of the intended application (radiation from a semiconductor device).

The experimenter is thus faced with a large number of choices and a limited time in which to do their work. The needs of practical experimentation – such as the construction of bespoke circuit boards; the sourcing, ordering and delivery of parts; the need to provide mechanical support to very fragile components; the need to reverse engineer components without available data; the need to fault-find and repair systems damaged in the course of experimentation; and the assembly, connection and configuration of a disparate collection of test equipment – mean that such evaluation is a time-consuming endeavour. In addition, the test equipment was shared: some was only available for a limited range of experiments, and when a new sensor was tested it was sometimes not possible to repeat experiments using the previous equipment. On the other hand newer equipment was used when it was purchased, which improved the quality of results at the expense of them not being directly comparable with previous experiments.

In this work, I performed brief testing of a wide range of potential sensors. Those that provided a glimmer of promise were further investigated. Those that showed no promise were discarded. Such experiments do not prove that these sensors are definitively unsuitable, but that the configuration in which I tested them did not show any results worthy of further investigation in the limited time available. Many of the constraints I outline above could be improved with further work.

4.2 ELECTRIC FIELD SENSORS

The simplest EM field sensors are the antennas which generally measure the electric field component of an EM signal. Most antennas are designed for specific frequency ranges and are poorly suited for broadband applications as required for EMA. A number of antenna topologies exist for broadband signals (some examples are given in Kraus and Marhefka

Figure 4.1: *Electric field dipole formed from stripped coaxial cable*

(2001, ch. 12)) but many antennas of a suitable size to be placed near a chip have their frequency band in the gigahertz region or higher, making them unsuitable for EMA of megahertz-band emissions.

A traditional passive antenna is designed to resonate at the desired frequency of reception and to match the impedance of free space to that of a coaxial cable. Instead, as suggested by Kuhn (2003), I tried an active antenna, an antenna with built-in amplification where matching is performed by the amplifier, the simplest of which is the dipole. A simple dipole was constructed by baring 16mm of core of RG58A/U coaxial cable and folding back the same length of braid (Figure 4.1). The resulting resonant frequency as a half-wave dipole is given by v/λ, where v in coaxial cable is approximately 0.8 times the speed of light. Thus this configuration has a resonant frequency of 3.75 GHz. As a resonant circuit this is not useful for EMA detection in the megahertz range. However, by connecting it directly to the 1 MΩ input of a Tektronix TDS2024 oscilloscope with 272 mm of RG58A/U coaxial cable it becomes an active antenna. An RG58A cable has a capacitance of 100 pF m^{-1} (Premier Farnell 2005), so a 30 cm length can be treated as a capacitive load of about 30 pF. For a 30 MHz signal, λ = 10 m and so the cable has an impedance of 177 Ω at 30 MHz. There will be some loss along the cable as a result; the loss can be minimised by shortening the cable, but this makes arranging the equipment more difficult.

Scanning the die area of the LH77790B microcontroller, running with 25 MHz clock, produced no signals synchronised with the operation of the processor despite the TDS2024 input gain being turned up to its maximum of 2 mV div^{-1}. By touching this probe on the chip package (Figure 4.2 on the following page), it could detect signals emanating from bond wires

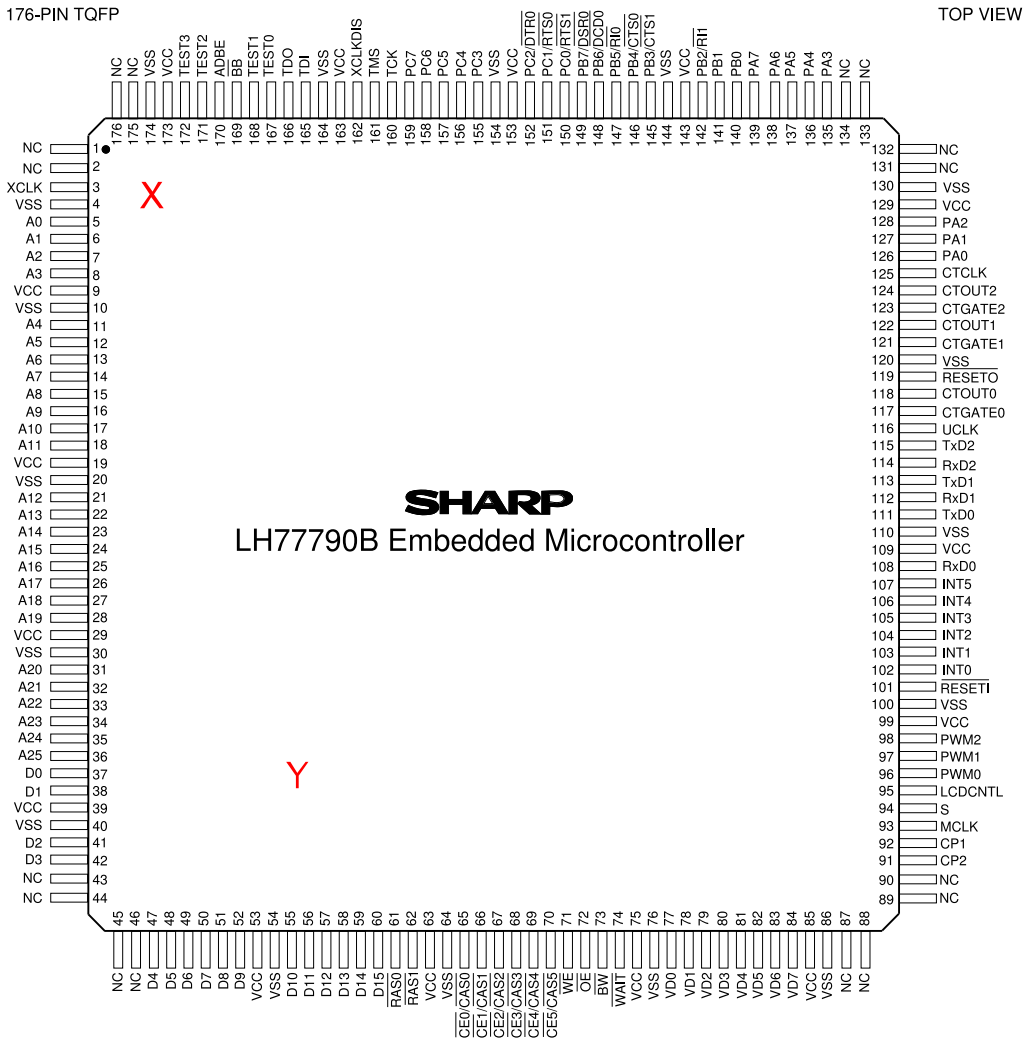


Figure 4.2: *LH77790B pin layout, reproduced from SHARP Microelectronics (1999). X and Y are electric field probe positions described in the following figures.*

carrying the clock (Figure 4.3 on the next page) and data bus (Figure 4.4 on page 82), but could not discern ALU operations of different Hamming weights. The difficulty with even a small dipole is that it is still comparatively large with respect to the area of the chip under test. At this point there was no spectrum analyser available for examination of its frequency behaviour.

Another broadband antenna is the capacitive sensor, as described by Smith (2003). Copper foil is applied to the upper surface of the chip's package, forming a capacitor between the die and the foil (Figure 4.1 on

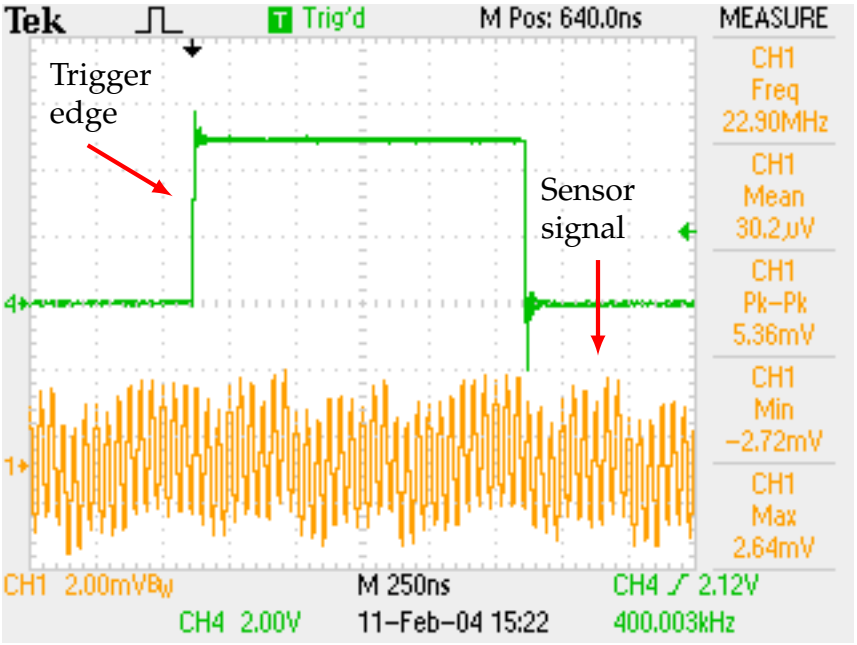


Figure 4.3: *End of half-wave dipole on LH77790B case orientation spot, near XCLK pin (marked as X on Figure 4.2 on the facing page). Appears to be detecting the 25 MHz clock frequency.*

page 79). Smith indicates that if a user holds the probe there is a parasitic path from probe body, through the user's hand to ground, which is sufficient to provide a return path at high frequencies to observe pickup. This is not repeatable for measurements, so it is better to connect a differential probe with one end on the foil and one end at a nearby ground.

This sensor was used for later measurements, reported in Section 4.7 on page 115 onwards.

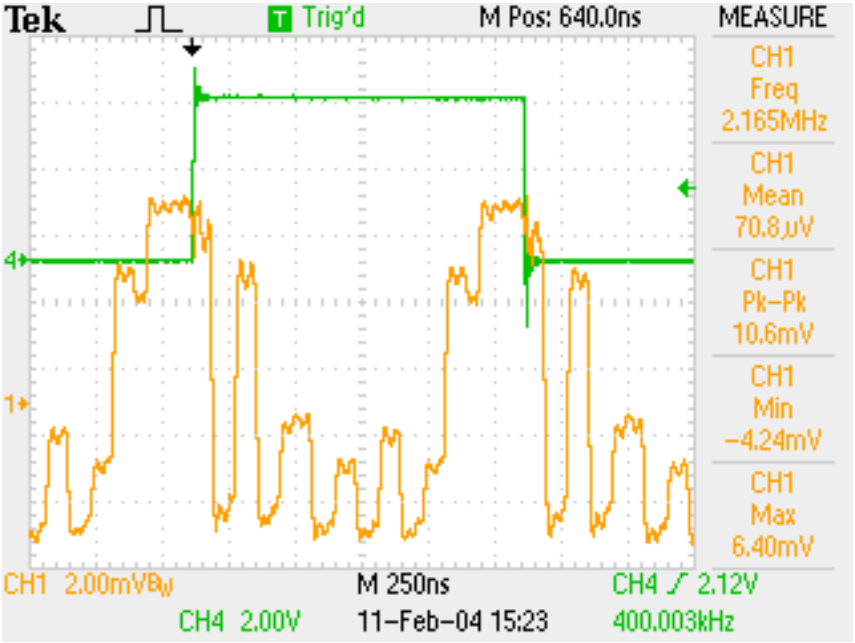


Figure 4.4: *End of half-wave dipole 5 mm in from LH77790B pin 55 (data bus area), marked as Y on Figure 4.2 on page 80. As would be expected, the trace shows much more variation.*

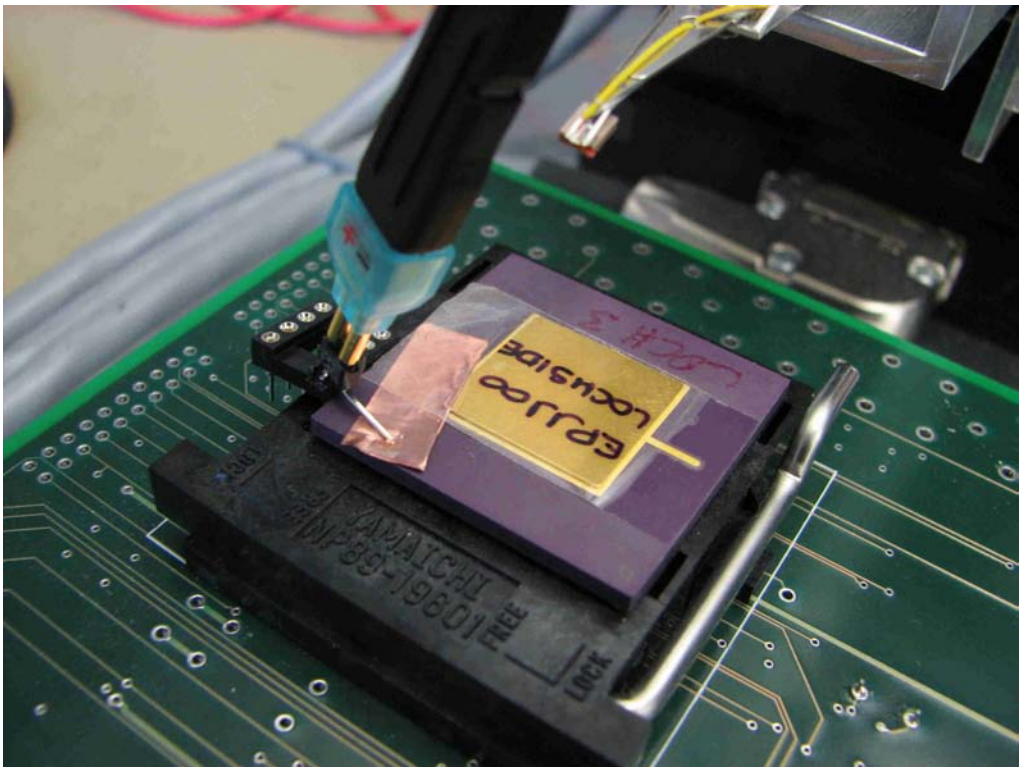


Figure 4.5: *Electric field sensor using brass chip cover of Lochside chip. The positive terminal of the P7330A differential probe is connected to the ground plane of the printed circuit board (PCB), while the copper foil connects the negative terminal of the probe to the brass die cover. The arrangement with the dual-in-line IC socket and oversized copper foil is so that it can be easily removed to allow other measurements (eg magnetic probing using the hard drive head in the background) and it will stick reliably to the ceramic package when in place.*

4.3 MAGNETIC FIELD SENSORS

In the near field, the magnetic field **B** at distance **R** emitted by a wire element **ΔL** carrying a current I is given by the Biot-Savart law (Section 3.1.6.2 on page 51).

There are two basic means of measurement: inductive sensors and direct-field sensors. Inductive sensors make use of Maxwell's version of Faraday's law of electromagnetic induction (reproduced from equation 3.13):

$$\mathcal{V} = \oint_L (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{L} - \iint_S \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S} \tag{4.1}$$

This relates the voltage $\mathcal{V}$ induced on a loop of perimeter **L** and surface **S** to both motion of the loop (first term above) and a changing magnetic field **B** (second term). We can ignore the first term since the sensor is fixed. A much simplified version is the transformer equation:

$$V = M\frac{dI}{dt} \tag{4.2}$$

where all the physical factors are rolled into M, the mutual inductance between the emitting circuit and the receiving circuit. As the induced voltage is dependent on the time derivative of the **B** field, the signal received by a sensing coil is proportional to the time-derivative of the current flowing in the circuit.

Direct-field sensors make use of some other effect, such as the Hall Effect or the Magnetoresistive Effect, to sense the magnetic field directly. Hall Effect sensors are those that use the Lorentz force (reproduced from equation 3.3):

$$F = q[\mathbf{E} + (\mathbf{v} \times \mathbf{B})] \tag{4.3}$$

for electric field **E** and magnetic field **B** where electrons drift in direction **v**. This force deflects a current I flowing in aconducting strip of thickness d. Thus the voltage across the strip is given by:

$$V_H = \frac{IB}{\rho d} \tag{4.4}$$

where B is the component of **B** normal to the strip and ρ is the resistivity of the strip.

The key here is the return of a voltage proportional to the **B** field, not a time-derivative as in the inductive case. The same principle applies to other effects, such as Anisotropic Magnetoresistance (AMR) and Giant

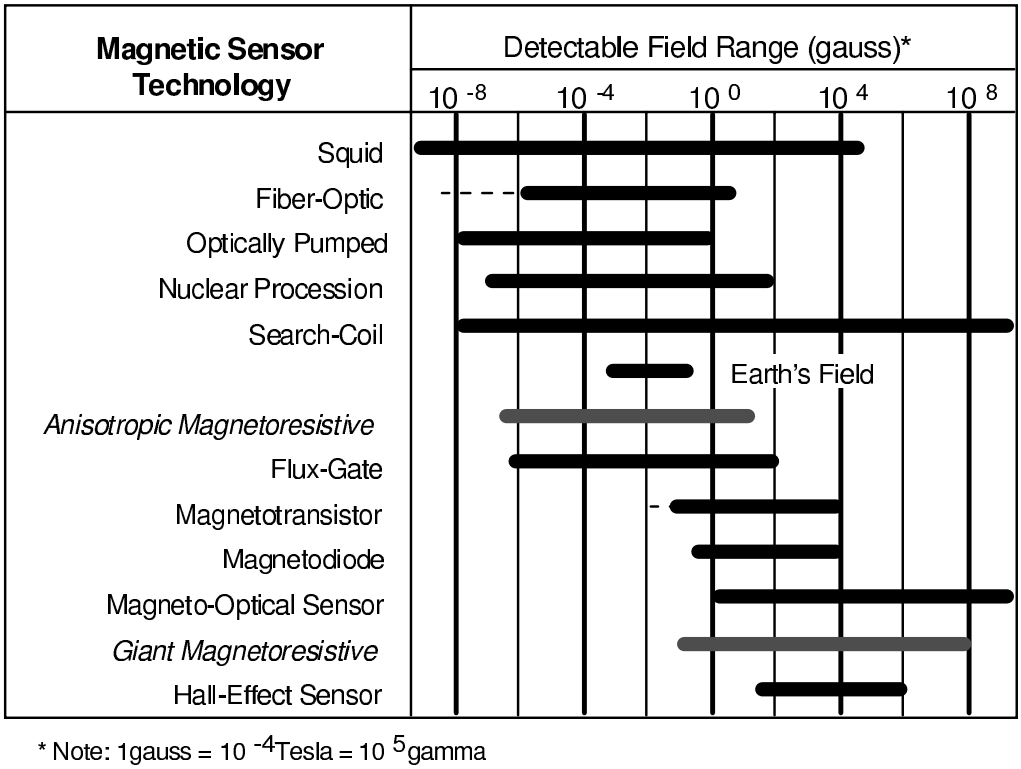


Figure 4.6: *Magnetic sensor technology field ranges (from Caruso, Bratland, Smith, and Schneider (1998))*

Magnetoresistance (GMR).

Figure 4.6 shows the relative sensitivities of various magnetic field sensors at their optimal operating conditions (temperature, frequency, etc). Note that not all sensors have conditions that are suitable for EMA applications.

The simplest magnetic field sensor is the coil. Since I was interested in targeted EMA, focusing on spatial resolution, I used prefabricated coils in the form of inductors and inductive hard drive heads which were made smaller than I could have wound myself. A small coil has a low output amplitude since it integrates fewer field lines, but I attempted to make up for this with a high level of amplification.

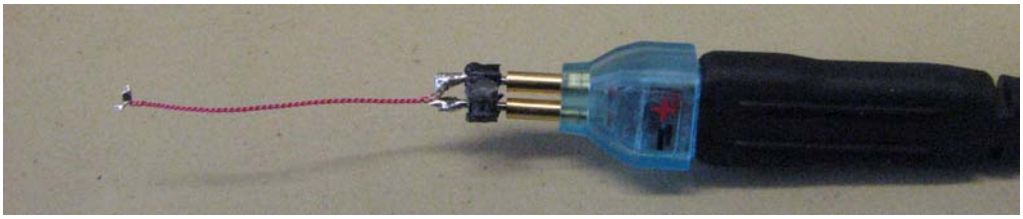


Figure 4.7: *1 nH 0402 inductor used as magnetic field probe. (For scale, pins in centre of picture are 2.5 mm apart.)*

4.3.1 1 nH 0402 inductor

A small probe was constructed: an 0402 1 nH inductor[1] (Tyco Electronics 36401E1N0A) on a self-made twisted pair cable, wired directly to the P7330A differential probe and TDS7254B oscilloscope (Figure 4.7). When an RF choke was connected to a signal generator, driven at 1 to 10 MHz at 5 V pk–pk, and pointed at different parts of the sensor, the gap in the twisted pair for the differential probe appeared to pick up more than the inductor.

4.3.2 Hard drive heads

The magnetic disc recording medium was first developed by IBM in the 1950s (Allan 2001). The company launched the first fixed Winchester disc for low-cost computer systems in 1973.

All subsequent drives follow the same basic pattern (Figure 4.8 on the facing page). A stack of magnetic discs (*platters*) are attached to a central *spindle*. The spindle is spun at a constant high speed by a motor. The *head arm* reaches across one side of the platters from the circumference to the centre. The head arm swings across by either a stepper motor or a voice coil. On the end of the head arm, above and below each platter, floats the *head* on a small air gap created by the aerodynamics of the head and platter. Each head is composed of a sensing element (either a coil or some other kind of magnetic sensor), and a *slider*, a block of material that optimises the head's aerodynamic behaviour. The mechanism is sealed by a lid with a small filtered breather hole to equalise pressure with the exterior.

[1] 0402 is a surface mount package size: it describes a rectangular two-terminal component having dimensions 0.04 in × 0.02 in (approximately 1 mm × 0.5 mm). Such components are typically used in mobile phones where space is at a premium. Larger sizes in the same range are 0603, 0805 and 1206.

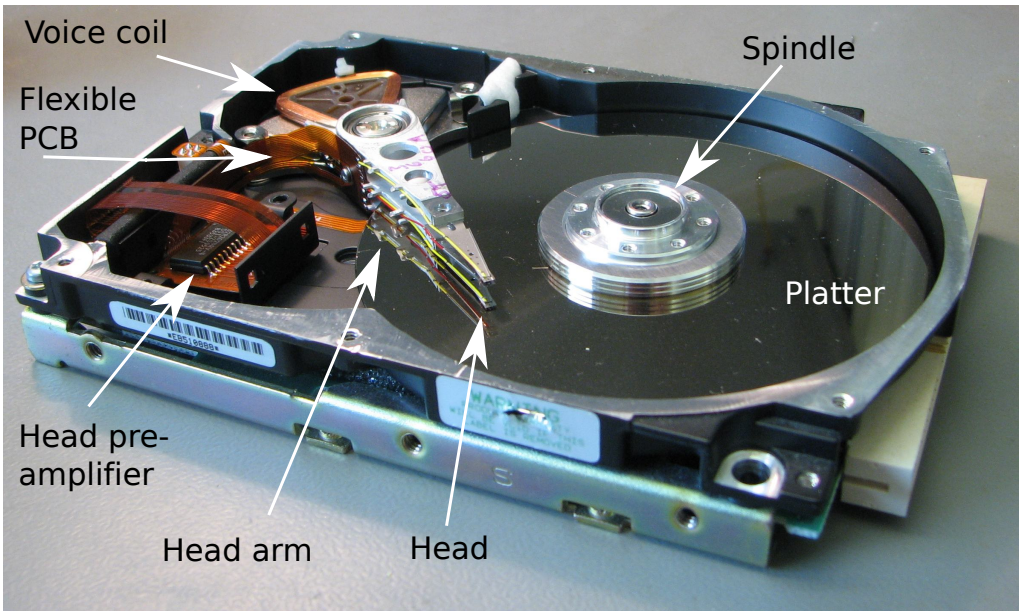


Figure 4.8: *Anatomy of a hard drive. One platter has been removed to reveal the positions of the heads.*

The magnetic element of each head may be composed of two parts – read and write – which are typically optimised for those functions and may be of different technologies. Each head is typically connected to two, three or four wires, twisted together. These wires run down the head arm to a *head preamplifier*. In older drives the head preamplifier is a surface-mount IC mounted on a static flexible PCB (*head flexi*) underneath the top cover; in more recent drives the IC may be mounted on a flexible PCB on the head arm itself and flip-chip bonding may be used.

Storage devices have been a primary driver of magnetic sensors in recent years with the need for ever-growing storage densities. Despite shrinkage, the write head has typically remained inductive throughout, while the read head has progressed through several technologies since 1990. Very roughly these developments are:

Pre-1990	Inductive ferrite cored
1990-1995	Thin film inductive
1995-1999	Anisotropic magnetoresistive (AMR)
1999-2001	Giant magnetoresistive (GMR)
2001-2003	GMR or GMR variants (spin valve, spin dependent tunnelling)
2004-	Collossal magnetoresistive (CMR)
2005-	Tunnelling magnetoresistive (TMR)

The difficulty facing anyone wishing to use hard drive heads for EMA is that all the devices within the lid are usually proprietary, with no public documentation. Sometimes a small amount of data for the head preamplifier chip is available, but typically it is necessary to guess the likely head technology and connections.

4.4 INDUCTIVE SENSORS

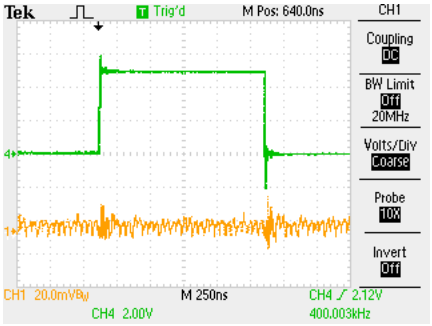
4.4.1 WDI325Q 20MB MFM hard drive head

The head from a 20MB ST506/MFM[2] hard drive made by IBM in the late 1980s, model WDI325Q, was tried as a magnetic sensor over the LH77790B microprocessor. This is an inductive head, with a large slider around it. Two heads were connected to the Tektronix TDS2024 oscilloscope, and compared. One had the coil mounted in the full head slider, the other had the slider broken off so that a small amount of core material remained inside the coil. In both cases, one winding of the two-winding coil was shorted out.

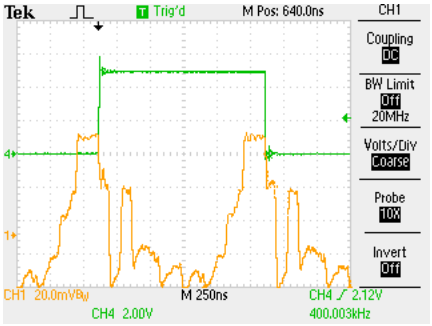
When positioned over the 'L' on the LH77790B packaging (the position of the strongest signal detected), the measured signal was much stronger with the broken slider than with the full slider (Figure 4.9 on the next page).

If a ST506 hard drive has a data rate of 5 Mbit/s, under MFM encoding each bit becomes a flux reversal. ST506 drives read from one head at a time; unlike more recent drives they do not parallelise the data coming from multiple heads. With one bit being one flux reversal, and one flux reversal giving one transition of the electrical output signal, the frequency of the magnetic signal from each head coil during normal operation is approximately 2.5 MHz. The LH77790B microcontroller is clocked at 25 MHz, and so it is likely that the head slider increases the inductance, creating a low-pass filter and filtering out the signals of interest. A higher-frequency noise signal remains, but this may be leakthrough of the 25 MHz clock (Figure 4.3 on page 81) directly into the amplification circuit rather than into the sensor (no spectrum analyser or better oscilloscope

[2]ST506 is a 1980s hard drive interface standard: it is based on control signals to the drive mechanism plus a bitstream from the head. All control and decoding is performed on the host adapter card. It is based on the 1976 Shugart SA400 interface to floppy drives used in most PCs. Modified Frequency Modulation is a means of encoding data onto the bitstream; the other common format is Run-Length Limited (RLL) encoding which has superseded MFM.



(a) WDI325Q head output with full head slider



(b) WDI325Q head output with broken head slider

Figure 4.9: *WDI325Q inductive head over LH77790B microprocessor measuring I/O transients caused by a trigger signal generated by the chip. The large magnetic 'mass' of the head slider reduces the head response.*

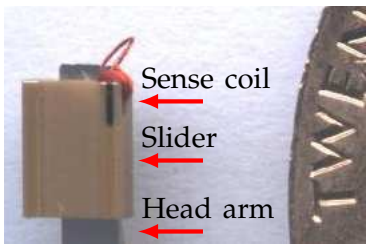


Figure 4.10: *Inductive head from Western Digital AC280 hard drive, circa 1990 (20 pence coin for scale)*

was available to investigate further at this point). All of the figures for the LH77790B I include here (using any sensor) show some kind of background circa 25 MHz signal. In any case, no data-dependencies could be detected from the LH77790B.

4.4.2 AC280 80 MB inductive hard drive head

Since the WDI325Q head was not successful, a more recent inductive head was tried. This was taken from a Western Digital Caviar AC280 drive, dated 1990, holding 80 MB (Figure 4.10).

The AC280 head outputs four wires, which are joined to three terminals on the head arm. One terminal takes two wires twisted together, so I infer that there are two coils and this terminal is the centre tap. The para-

Figure 4.11: *AC280 head L/R parameters measured on LCR bridge. To the left are shown the parameters from measuring each head coil separately; on the right, from measuring across both coils in series. The discrepancy in inductance is likely to be parasitic inductance from the LCR connections.*

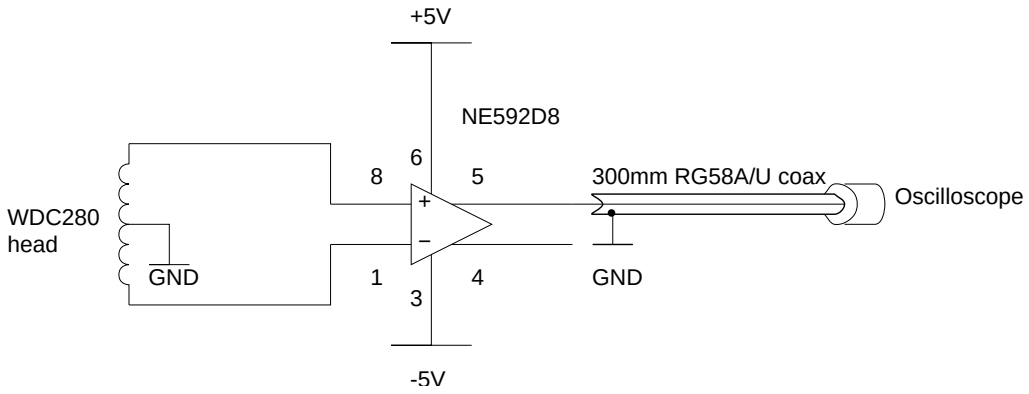


Figure 4.12: *Test circuit for AC280 inductive head*

meters as measured by a Tinsley Prism 6451 LCR[3] bridge may be seen in Figure 4.11.

Initially, the centre tap was grounded and the end of each coil was wired to an NE592 differential amplifier chip with a fixed gain of 400, as seen in Figure 4.12.

When positioned in the same place over the LH77790B databus as shown in Figure 4.4 on page 82, the probe did detect a magnetic signal as seen in Figure 4.13 on the next page. This shows its inductive nature, by producing impulses that look similar to the differential of Figure 4.4 on page 82 (but not verifiably so).

A timing attack was tried using the one-instruction `loop1` and two-instruction `loop2`, which may be compared in Figure 4.14 on page 92.

[3]Inductance, capacitance, resistance

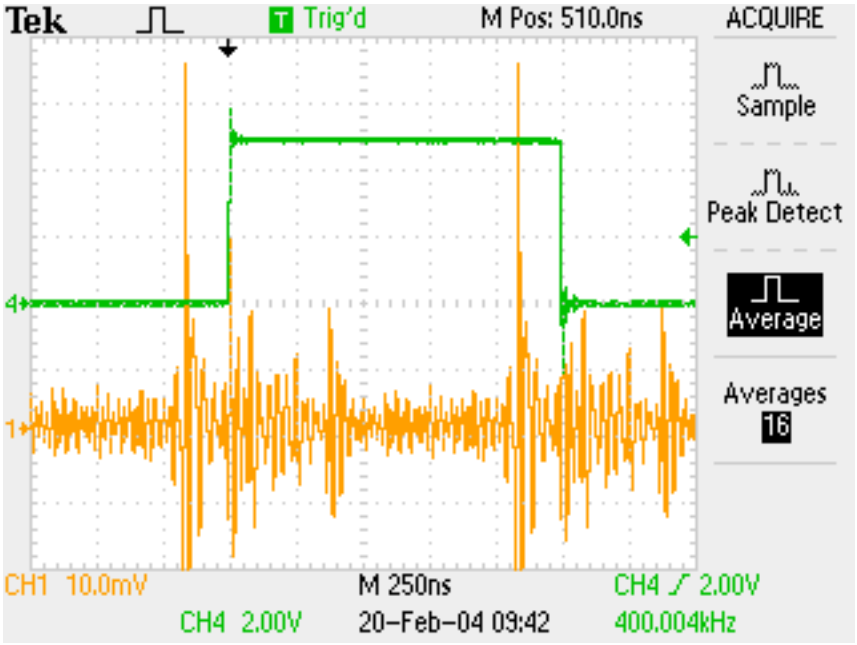
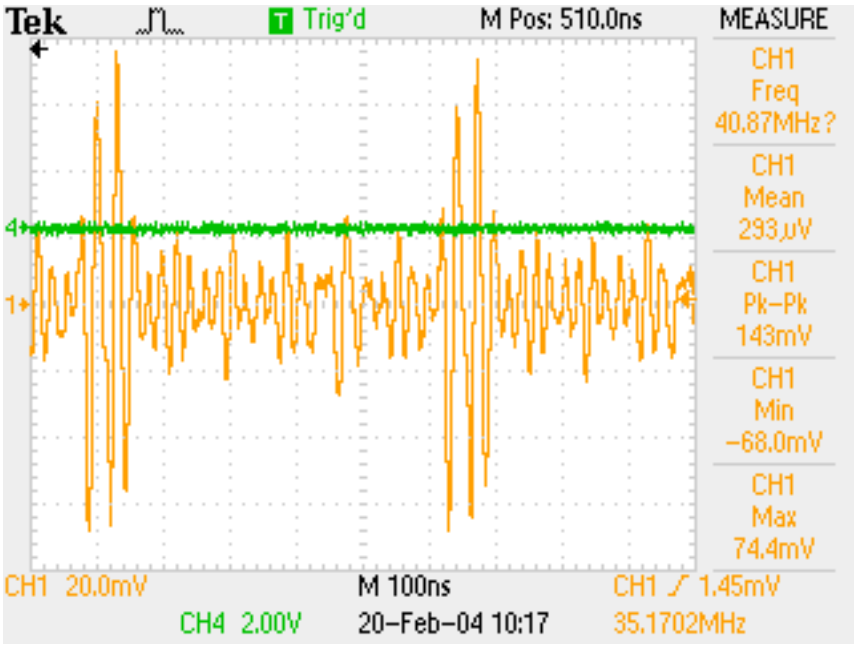
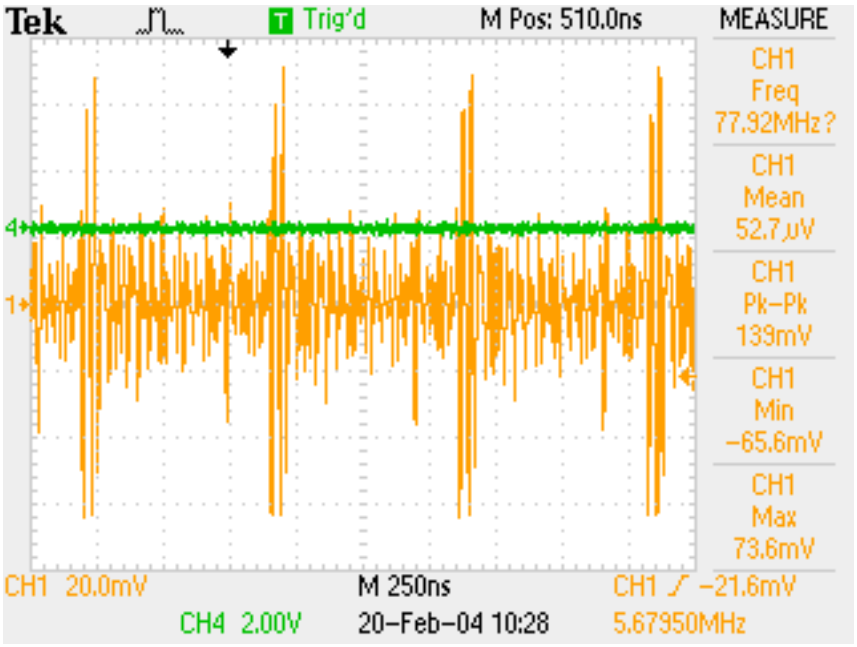


Figure 4.13: *AC280 inductive hard disc head 5 mm in from LH77790B pin 55 (data bus area), in same position as Figure 4.4 on page 82. There is obviously some pickup of a signal synchronised with the microcontroller clock. While it is not clearly the differential of the electric field signal, the impulsive nature of the inductive sensor resulting from it measuring dI/dt can be seen.*

The performance of the AC280 was interesting enough to build a better probe. A revised circuit using lower-noise amplifiers and a precision voltage reference was designed and built by Sergei Skorobogatov. This may be seen in Figure 4.15 on page 93. This board was used for the experiments using the spectrum analyser and 3D spatial scanning as described later in this chapter.



(a)



(b)

Figure 4.14: *Timing analysis with AC280 head on LH77790B: (a) one-instruction loop; (b) two-instruction loop.*

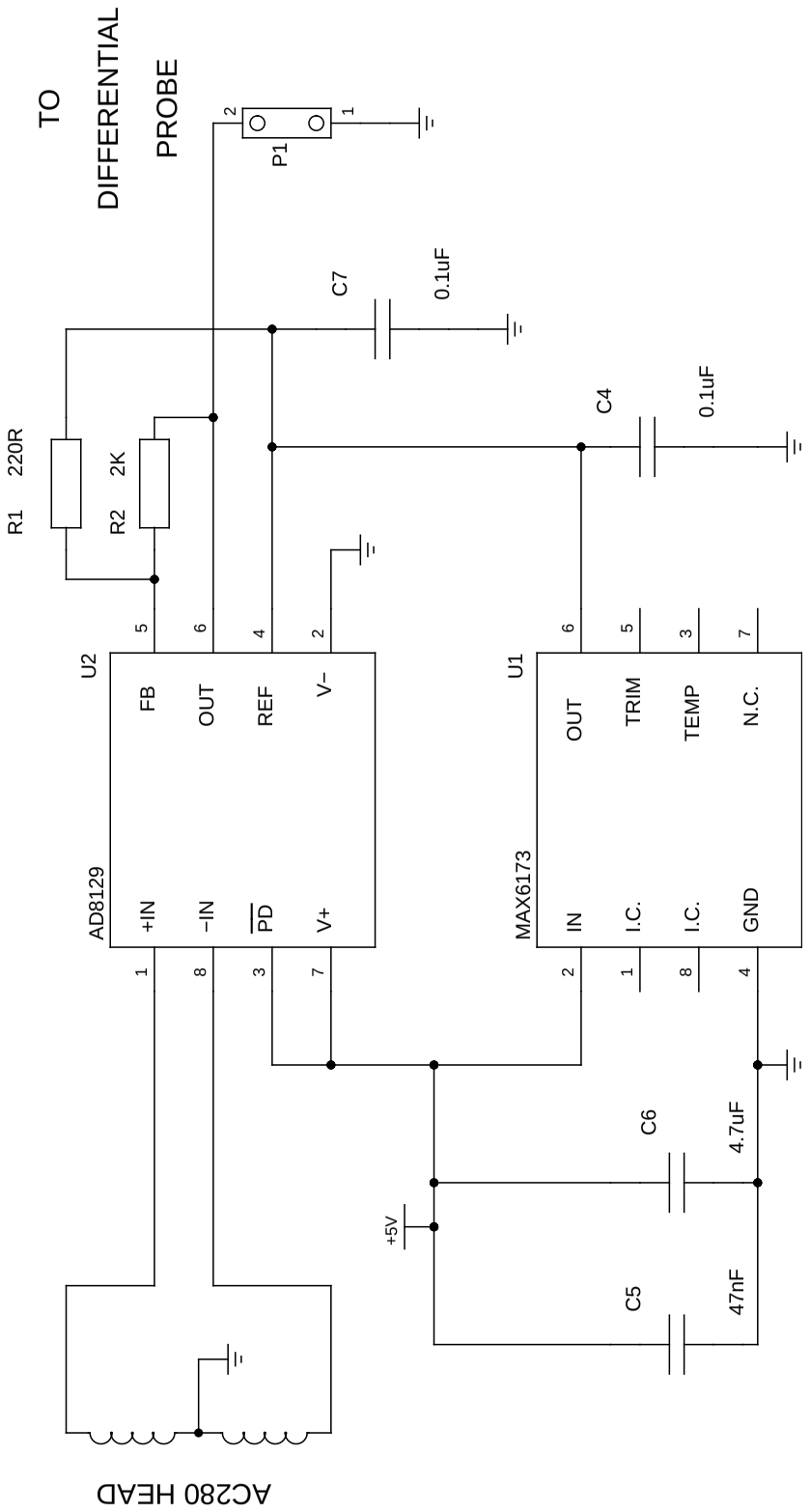


Figure 4.15: *Improved low-noise amplifier for AC280 inductive head*

4.4.3 Samsung 1 GB, inductive head

4.4.3.1 Using SSI 32R2212R head preamplifier

The head from an unidentified 1 GB Samsung drive[4] was tested, to see if the head preamplifier chip could be used as an amplifier matched to the head and driven in the way that it was designed. No data was available for the head or preamplifier; the preamplifier was an SSI 32R2212RX4 from Silicon Systems Incorporation, latterly Texas Instruments.[5] On the mainboard the detector chip was the SSI 32P4910A.

At the time of these experiments, no data could be found for the 32R-2212R, though summary datasheets for chips in the 32R22xxR series were available and the chip is mentioned in patents (Yun 2001, Ahn 1999). The combination of preamplifier and detector used was mentioned in Cain, Payne, Qiu, Latev, Imai, Hempstead, McNeil, and Phenicie (1996) which indicates the signal conditioning chain is designed for an inductive sensor. That publication, which predates the hard drive by some months, also implies the recording density is about 1 Gbit/in^2, which means a sensor with a bit dimension of $\sqrt{(0.0254^2)/10^9} = 800$ nm.

Platters in 3.5 in format drives have an inside diameter of 25 mm and outside diameter of 95 mm, which gives an area of 6600 mm^2 per side. The 1 GB=8 Gbit Samsung drive has four surfaces, giving a total area of 26 400 mm^2 or 40.9 in^2. Assuming this is all usable surface for data, that gives a recording density of 196 Mbit in^{-2} and thus a head size of 1.8 µm. The missing factor of five in the bit density is easily explained if not all the platter area is available, owing to servo tracks, data encoding requirements and private storage used by the controller for block remapping. Also this may not be an example of the newsworthy drive described in Cain, Payne, Qiu, Latev, Imai, Hempstead, McNeil, and Phenicie (1996). However, it means we can expect a head size of about 1 to 2 µm.

This is a considerably smaller size than has been achieved in other EMA work, such as the 40 µm sensor in Quisquater and Samyde (2001), and potentially allows us to achieve a much higher spatial resolution.

First, the drive casing was opened. The drive continued to work as normal. Hard drives typically do not have anti-tamper switches to prevent them functioning with the casing removed; indeed the AC280 drive would reliably boot Windows 3.1 for several weeks after its lid had been

[4]Drive marked WINNER REV. E and dated 97-07 (most likely March 1997). It may be from the same family as the WINNER-1 WN31601A 1.6 GB drive (Samsung Electronics 1997)

[5]Package labelled as 2212RX4, also referred to as a 2212R or 32R2212R

removed and its platters were collecting dust. The Samsung heads were removed for inspection, then replaced. The drive spun up twice, then shut down. Subsequently it would not spin up again, despite power cycling and checking the connections. I concluded that the drive has some internal protection mechanism which means, in the case of an error, it will shut down permanently to protect data on the platters. One datasheet I found (Atmel Corporation (2004), for the Atmel AT78C6002 which I didn't have available to test) does include a fault detection circuit that raises an alarm to the drive mainboard.

This meant that it was not possible to monitor data being streamed off the heads when in operation on the platters. Due to the lack of a datasheet I was forced to reverse engineer the 32R2212R preamplifier chip circuit and its connections to the 32P9410 detector. The head wiring can be seen in Figure 4.16 on the next page, and the area on the mainboard it connects to in Figure 4.17 on page 97.

No data was available for the 32P4910A either, but it is mentioned in a patent (Takahashi and Kimura 2000). Looking at the data sheets for other pre-amplifiers gives us an idea of what signals to expect. For example, the VM7200 (VTC Inc. 1993) is an earlier pre-amplifier designed for thin film inductive heads. Most drives only read one head at a time and the pre-amplifier includes an analogue multiplexer. Our chip has four head input pairs, so we expect two head select logic wires. The read output signal is produced on another balanced pair. In addition to these, the VM7200 requires power and ground (easily discernible thick tracks on the flexible PCB), a chip select and a read/write signal. The Write Current Adjust pin has a resistor to ground to set the current. The Write Data Input (WDI) pin receives the digital data to be written in the form of TTL transitions, while the Write Unsafe pin gives a high when the disc is not available for writing (when in read mode, no write current or insufficient frequency on WDI).

Of the four pairs of traces from the 32R2212R, the pair with 51 Ω termination was guessed to be the analogue signal output from the preamplifier: each signal has a series capacitor before it enters the 32P4910A. Another pair goes directly to the 32P4910A, so these are assumed to be logic level control outputs, perhaps head select. There are three other signals with series 100 Ω resistors, which are also assumed to be logic signals, perhaps high speed. On the mainboard, one pin is connected to ground via a precision 2.67 kΩ resistor, so this is assumed to set the write current. My connection guesses may be found in Table 4.1 on page 98.

The six unknown connections A to F were wired with 100 kΩ pull-up resistors to 5 V and DIP switches to pull them low. A 2 kΩ resistor to ground was wired on WCS to set the write current (although I am only

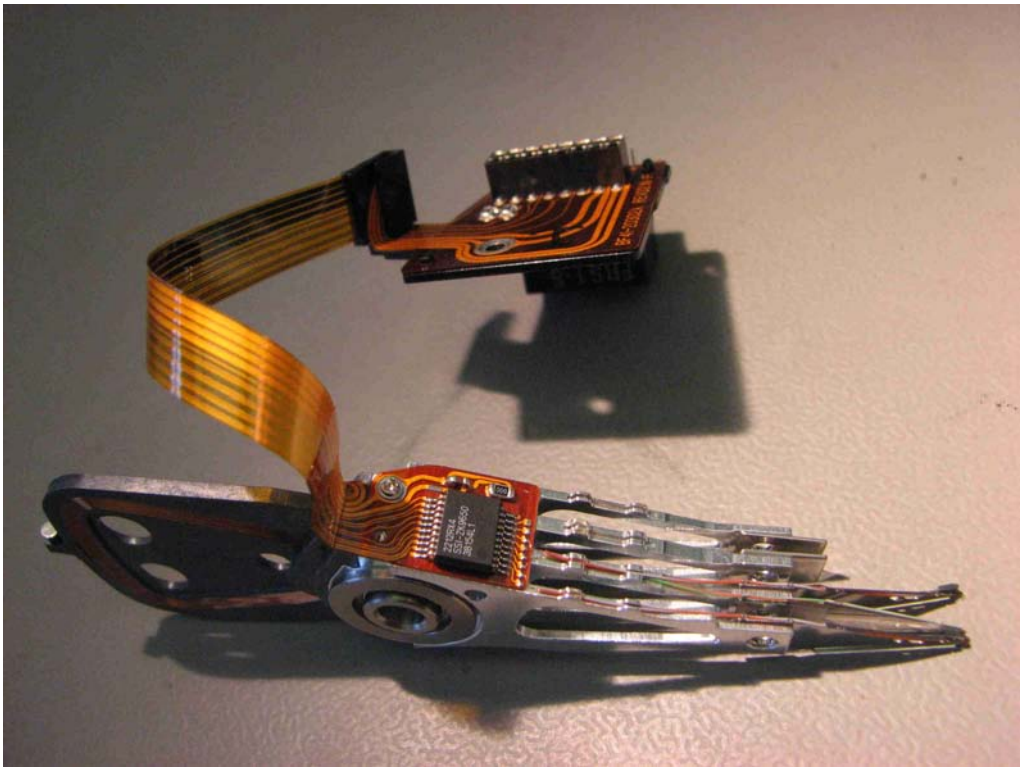


Figure 4.16: *Head assembly of Samsung 1 GB drive showing heads (one missing), head arm, arm-mounted 32R2212R preamplifier and flexible PCB to mainboard. Each head can be seen to have two wires connecting it to the preamplifier chip. The solder connections to the voice coil actuator are the two tracks to the left of the fixing screw, while above the chip are the power supply traces. The remaining connections consist of four pairs of traces and one single trace. One pair is terminated at the connector with 51 Ω resistors.*

interested in read mode). With some experimentation, the settings seen in Table 4.2 on page 99 were found to detect a 1 MHz magnetic field placed nearby (Figure 4.18 on page 99). Flipping signal F had a noticeable effect on both the supply current drawn and the output on the differential pair, which lead me to believe that is the Read/$\overline{\text{Write}}$ signal.

The hard disc platter turned out to be good magnetic screening for the heads, so it was first necessary to remove a platter to receive a good magnetic signal. Without any external magnetic field, there was no detectable output change when the platters were spun by hand. This suggests there may be a high pass filter present. Manually spinning at perhaps 60 rpm is approximately two orders of magnitude slower than normal speed of 4200 to 7200 rpm, and would scale the frequency appropriately. Alternat-

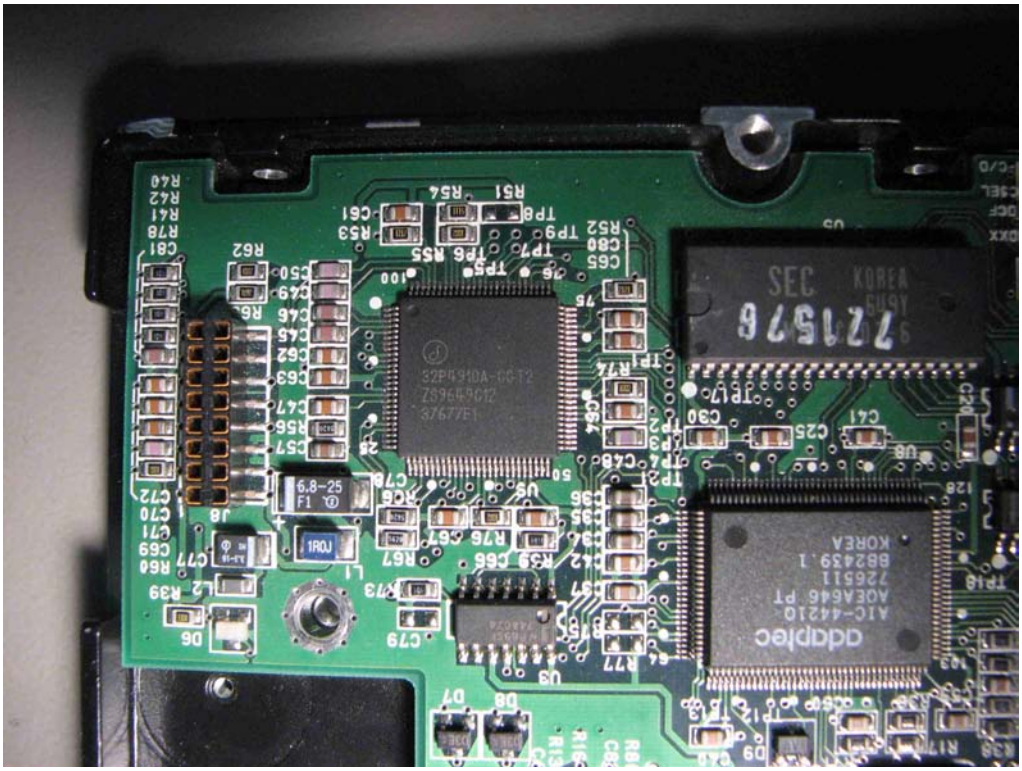


Figure 4.17: *Section of Samsung 1 GB mainboard with 16-pin connector to head assembly on left. Many passive components are used to interface the head chips to the SSI 32P4910A detector chip.*

ively the geometry is wrong. In operation the heads would normally float on a thin air cushion above the platter, here the heads are touching the surface due to the low spin velocity.

When the magnetic probe produced a 200 kHz field, it was not detected by the head and preamplifier chain. This supports the theory that there is high pass filtering. A 10 MHz field was received much better than a 1 MHz field. This may just be induction into the (twisted pair) head wiring, or it may be due to filtering. It is quite likely that the preamplifier includes a bandpass filter which is tuned to the data rate expected from the disc. Therefore I decided to try using the head on its own.

Many months after this work was completed, outline datasheets for the 32R2212R and 32P4910A were discovered (Texas Instruments Storage Products Group 1998a, Texas Instruments Storage Products Group 1998b). While some of my assumptions about the 32R2212R wiring were wrong (the head select pair are pins 14/15, not 12/13), the inputs in Table 4.2 on page 99 are correct for read mode. Respectively they consist of Servo Bank

32R2212R pin	Flexible connector pin	Guessed function
	1,2	Voice coil
1	3, 6, 11, 20	GND
2, 3		Head pair 0
4, 5		Head pair 1
6, 7		Head pair 2
8, 9		Head pair 3
10	4	VCC
11	5	A?
12	8	Head select B?
13	10	Head select C?
14	7	Digital data D?
15	12	Digital data E?
16	14	Read data X?
17	16	Read data Y?
18	15	Write current set (WCS)?
19	13	F?
	9	Not connected

Table 4.1: *Samsung 1 GB drive head wiring. Horizontal lines keep together pairs on the flexible PCB.*

Write (WUS)=Low; Write Data=High; $\overline{\text{Write Data}}$, ($\overline{\text{WD}}$)=High[6]; Head Select 1 (HS1)=Low; Head Select 0 (HS0)=High; and Read/$\overline{\text{Write}}$ (R/$\overline{\text{W}}$) = High. Driving head 1 may have been the reason it was necessary to remove a platter. In Figure 4.18 on the next page the platter between heads 2 and 3 has been removed, so it no longer screens heads 1 and 2. While the magnetic probe was targeted at head 3, the magnetic field would have been broad enough to cover heads 1 and 2 as well.

4.4.3.2 Samsung 1 GB head directly connected

A single head was removed from the head arm and connected to the circuit shown in Figure 4.19 on page 100.

[6]The conflict Write Data=High with $\overline{\text{Write Data}}$=High should not matter since the preamplifier is in read mode where these signals are most likely ignored.

Pin name	32R2212R pin number	Logic level
A	11	Low
B	12	High
C	13	High
D	14	Low
E	15	High
F (R/$\overline{W}$?)	19	High

Table 4.2: *Samsung 1 GB inputs to enable magnetic detection*

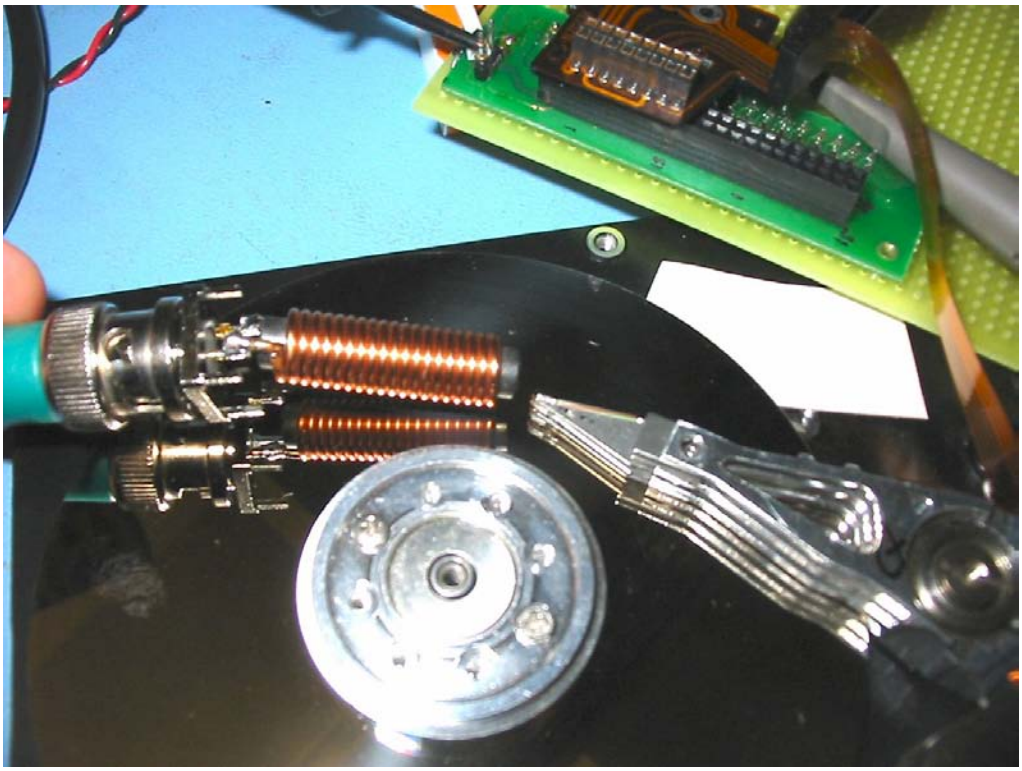


Figure 4.18: *Testing Samsung 1 GB head using integral head preamplifier (seen on test harness in top right corner)*

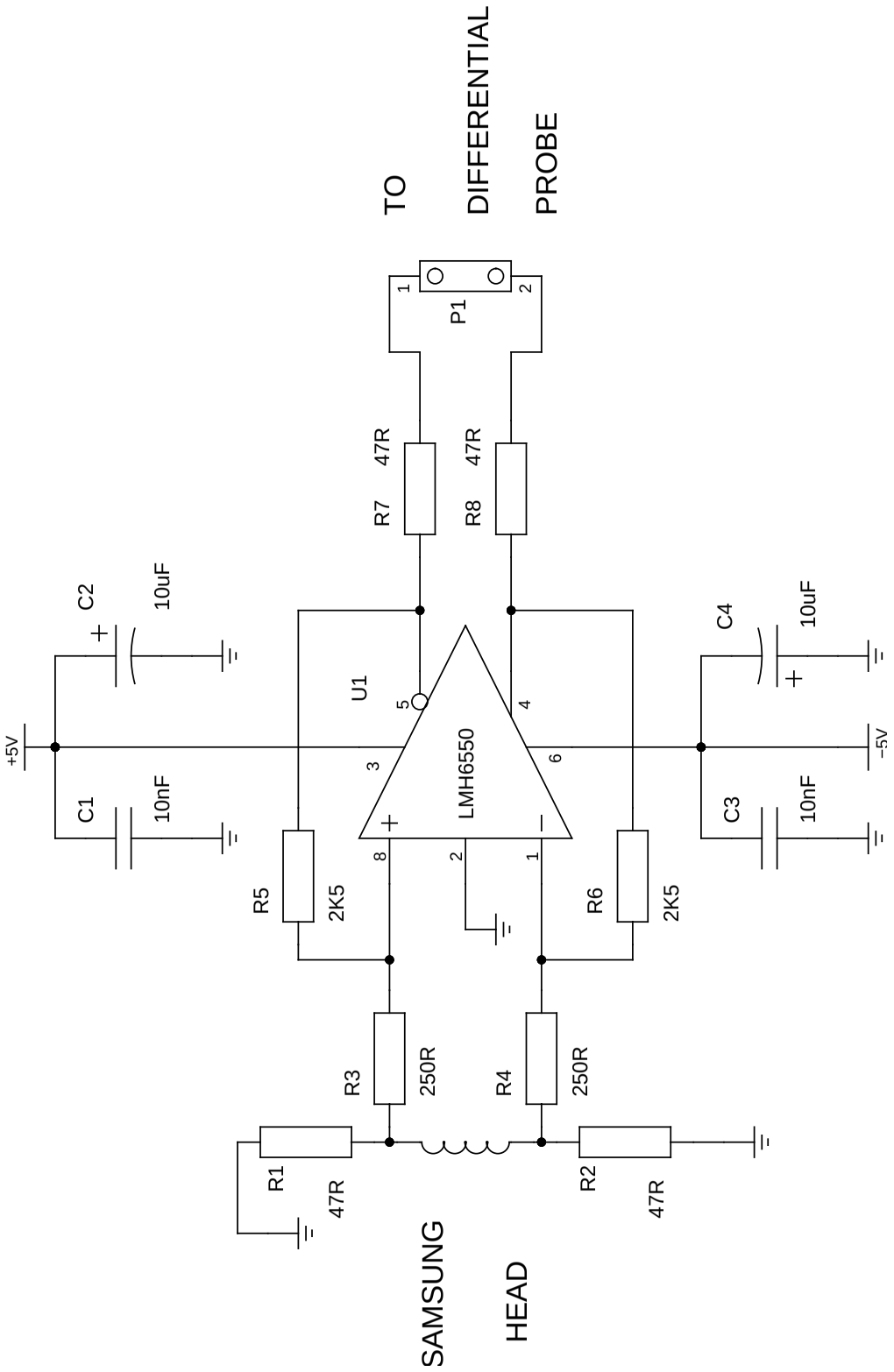


Figure 4.19: *Amplifier directly connected to Samsung 1 GB head*

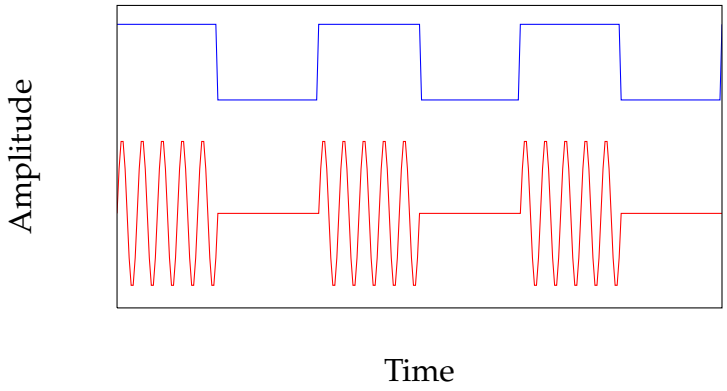


Figure 4.20: *Amplitude Shift Keying. By triggering the oscilloscope on an edge of the blue keying square wave, we can look for the carrier signal present on one side of the edge but not on the other. This makes it easier to distinguish the ASK signal from a harmonic of interference from elsewhere.*

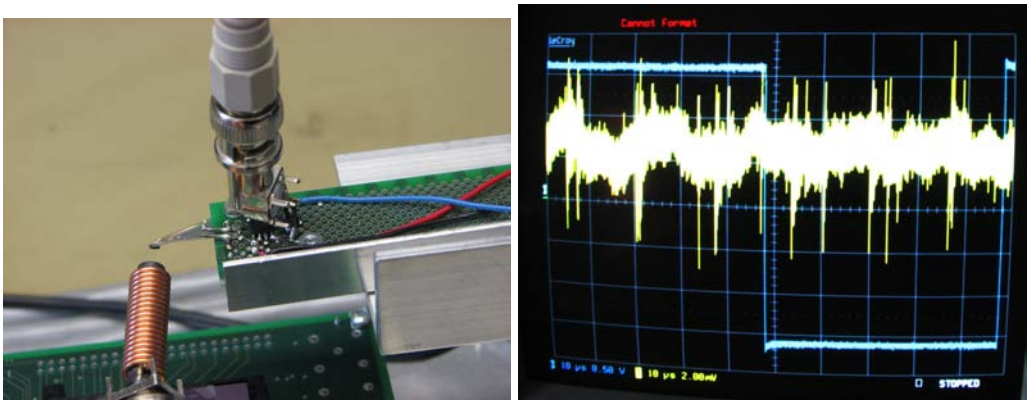
To more accurately ensure the signal being measured was the magnetic signal, the magnetic signal was a 1 MHz 5 Vpk–pk sinusoid 100 % amplitude shift keyed (ASK) with a 10 kHz square wave. By triggering on an edge of the 10 kHz keying signal, we can look for the transition between 1 MHz detectable and non-detectable. An example illustration of the ASK input to the magnetic coil may be seen in Figure 4.20.

The ASK results (Figure 4.21 on the following page) indicate minimal reception by the head, but a much stronger reception of magnetic field by the amplifier circuit. While the amplifier could be shielded, the lack of reception by the head of even such a strong field means it looks doubtful as a potential sensor. The same board was tried, connected to the Tektronix TDS7254B via a P7330 differential probe with similar results.

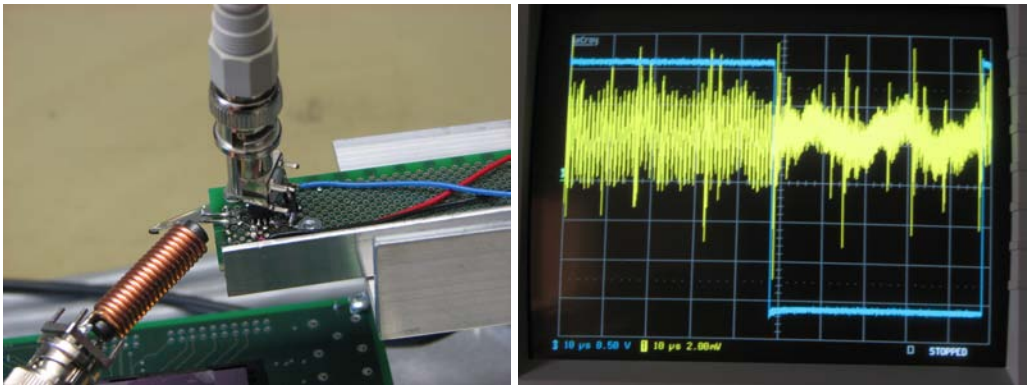
4.5 MAGNETORESISTIVE SENSORS

4.5.1 Anisotropic magnetoresistive (AMR)

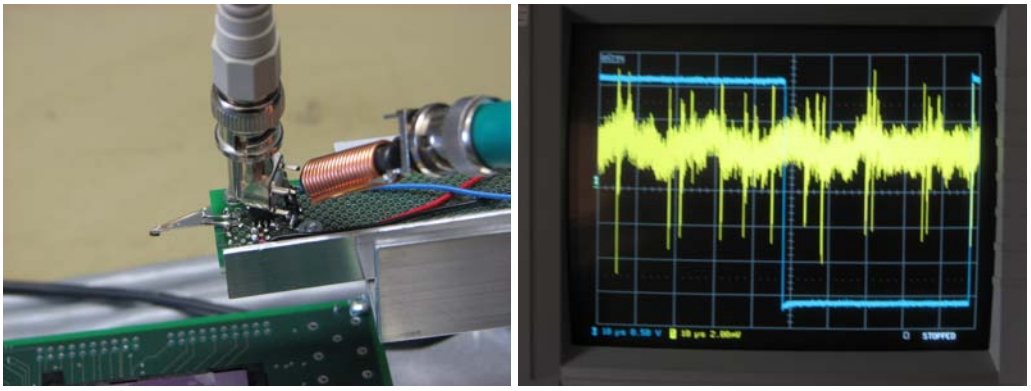
In the early 1990s hard drives moved away from inductive to magnetoresistive heads. Magnetoresistance was first discovered by Lord Kelvin in iron and nickel. He noticed that the resistance increases for a magnetic field parallel to the applied current and decreases when the magnetic field is perpendicular to the current. In metals magnetoresistance (MR) is secondary to the Hall effect, while in semiconductors MR is a first-order effect due to the Maxwellian distribution of electron velocities (Lark-Horovitz



(a) Pointing across head, end field lines linking into head coil



(b) Pointing at amplifier chip



(c) Pointing at coaxial connection

Figure 4.21: *Response of Samsung 1 GB head and amplifier board when magnetic field source applied in different places. 1 MHz frequency keyed by 10 kHz blue square wave: when high, 1 MHz magnetic field is produced; when low, it is switched off. Yellow trace shows output from head amplifier board. Only when the field is directed at the amplifier chip is a clear 1 MHz signal received, suggesting poor sensitivity of the head in this configuration. Results obtained using a LeCroy LC564A 1 GHz oscilloscope.*

and Johnson 1959). Both arise from the Lorentz force that acts on an electron moving in a conductor when a magnetic field is present.

The basic magnetoresistive effect is referred to as *Anisotropic Magnetoresistance* (AMR). Honeywell sell a variety of commercial AMR magnetometer devices. The most sensitive, HMC1001 and HMC1002 (Honeywell International Inc 2008), claim a resolution of 27 microgauss (or 2.1 mA m^{-1}) at 10 Hz, with a typical magnetic bandwidth of 5 MHz. The HMC1002 contains two sensing dice at 90° to each other; this is interesting as it allows measurement of both the magnitude and direction of a magnetic field. In my tests its outputs were each buffered by a CLC417, then amplified differentially by an NE592D8, as shown in Figure 4.22 on the next page. The offset and set/reset straps, used for calibrating the DC output and resetting the sensor when exposed to strong (10 to 20 gauss) magnetic fields, were left unconnected. A photograph of the board may be seen in Figure 4.23 on page 105.

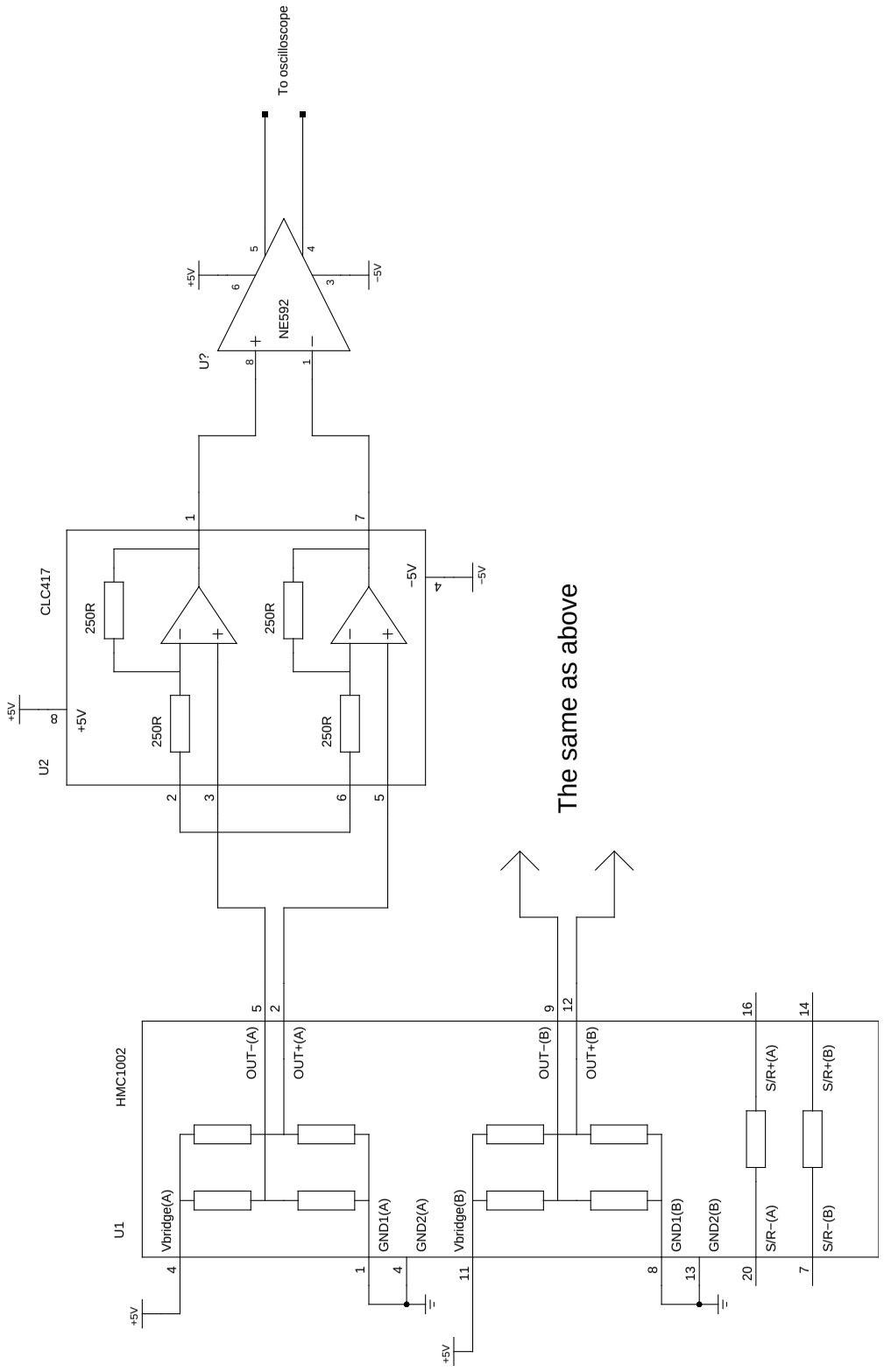


Figure 4.22: *Test circuit for HMC1002 anisotropic magnetoresistive sensor*

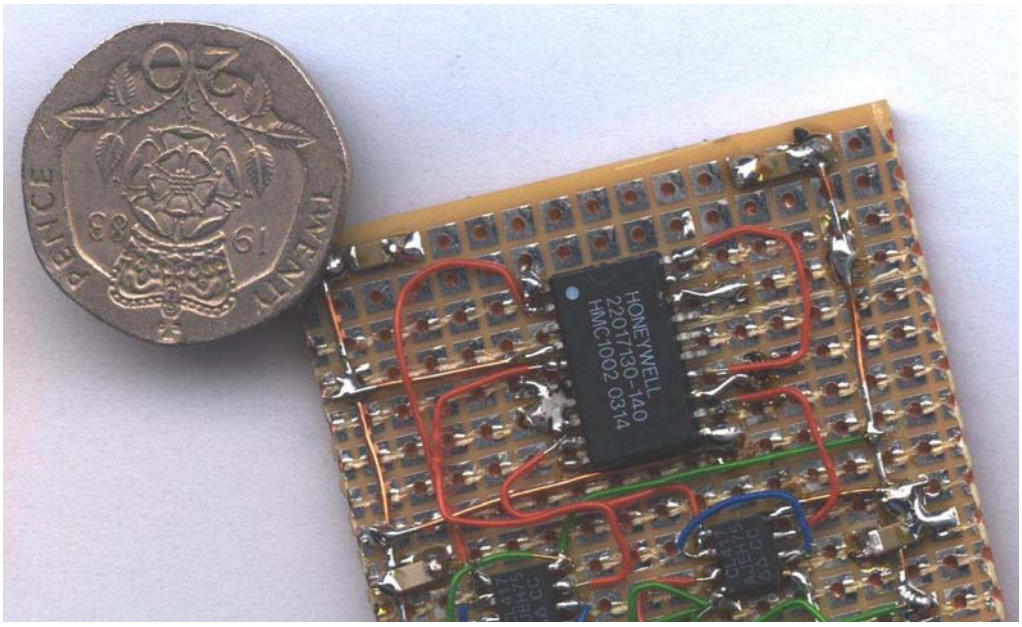


Figure 4.23: *HMC1002 anisotropic magnetoresistive sensor test board. The reverse side of the board is covered by a ground plane formed from copper tape.*

When positioned over the LH77790B, the dice detected the clock signal very well (lower trace of Figure 4.24 on the next page), along with some other detail. No ALU data-dependencies could be detected. This performance is notably better than the datasheet's suggested magnetic bandwidth of 5 MHz, though in addition there will always be some inductive or capacitive pickup from the board.

The HMC1002 datasheet does not reveal the size of the two dice encapsulated in the package nor their separation, but indicates they are side-by-side in the 13 mm long package. In a magnetic field with considerable spatial variation over this distance, the signals from the two dice cannot be combined to measure the localised field with any kind of accuracy. This can be seen in Figure 4.24 on the following page: the lower trace is further away from the active area, so it is just picking up the clock.

4.5.2 Giant magnetoresistive (GMR)

The Giant Magnetoresistive (GMR) effect was discovered in the late 1980s, and demonstrated large (up to 50%) resistance changes in thin layers of metallic elements (Baibich, Broto, Fert, Van Dau, Petroff, Etienne, Creuzet, Friederich, and Chazelas 1988). This required low temperatures and very

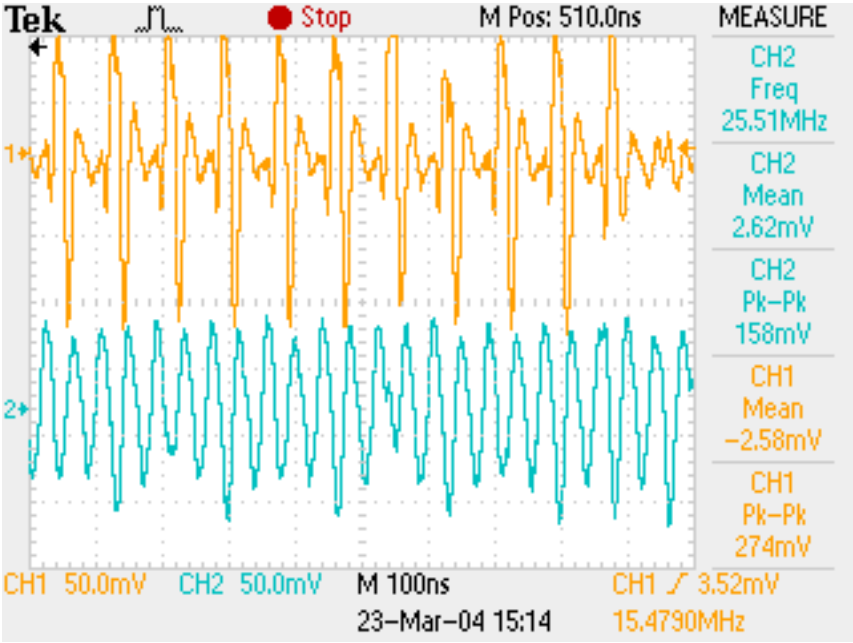


Figure 4.24: *AMR sensor over LH77790B running single instruction* `loop2` *test, two traces showing X (upper) and Y (lower) components of magnetic field. Note that the dice that measure components that are some millimetres apart.*

strong magnetic fields compared to the weak fields found in a hard drive.

The spin-valve magnetoresistor was published by IBM (Dieny, Speriosu, Metin, Parkin, Gurney, Baumgart, and Wilhoit 1991). Here a spacer layer of non-magnetic metal is sandwiched between two magnetic metals. When the spacer is thin, the two magnetic layers tend to align their fields.

In a spin-valve GMR sensor, the spacer is chosen thick enough so there is a weak coupling between the two magnetic metals. A fourth, strongly antiferromagnetic, layer is added to 'pin' the direction of one side of the sandwich. When a weak magnetic field is applied to the unpinned layer, its magnetic orientation rotates with respect to the pinned layer. The orientation difference gives a significant change in electrical resistance due to the giant magnetoresistive effect. This is the basis of modern GMR hard drive heads.

I tried a modern hard drive head using a giant magnetoresistive (GMR) sensor (Figure 4.27 on page 109). This came from an IBM Deskstar DTLA-307045, which holds 46.1 GB and was manufactured in May 2000. An IBM press release indicates:

> The IBM Deskstar 75GXP incorporates innovative new technologies to greatly enhance system performance – delivering an internal transfer rate that tops the industry average by 27 percent. Glass media, differential preamplifier, and fifth generation GMR heads enable this level of performance by increasing track capacity an average of 41 percent.
>
> (IBM Storage Technology Division 2000)

Each head provides four wires connected to a flip-chip head preamplifier mounted on the head arm. Being an unmarked flip-chip the head preamplifier was of limited use. For an individual head, the resistances can be seen in Figure 4.25 on the following page.

In the absence of other information, I assumed that the 36 Ω pair was the magnetoresistive element. By placing a 470 Ω resistor in series and applying 5 V, I set the bias current to 10 mA (a guess based on reading datasheets for MR head preamplifiers). I attached it to a preamplifier formed from a high impedance CLC417 buffer and gain 400 NE592D8 amplifier, as seen in Figure 4.28 on page 110.

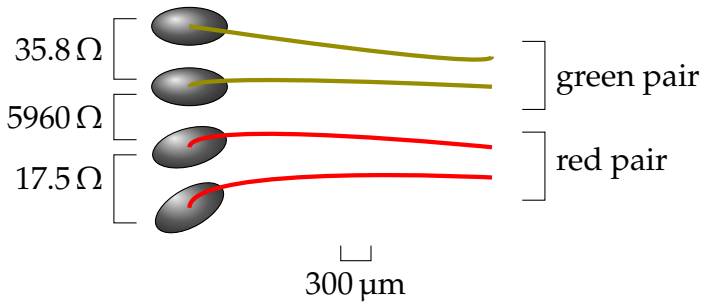


Figure 4.25: *Head resistances of IBM DTLA-307045 GMR head connections, at the junction between the head twisted pair wiring and the flexible PCB*

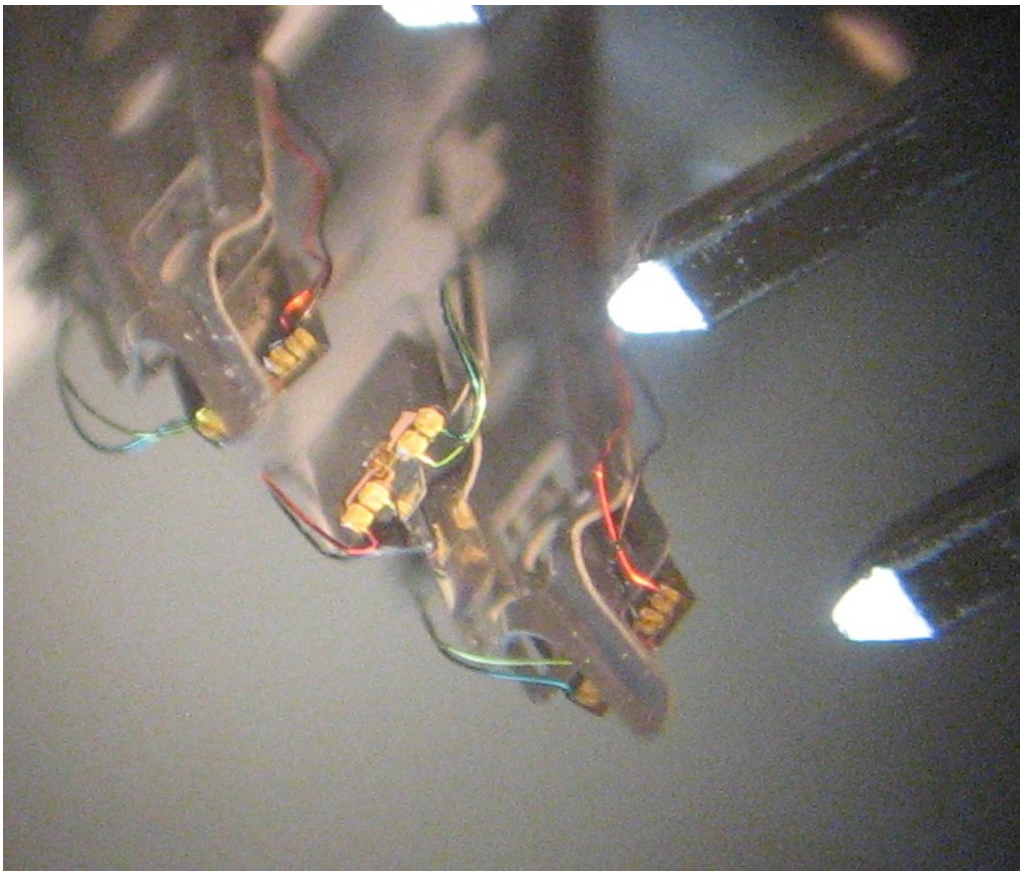


Figure 4.26: *GMR head from IBM DTLA-307045 drive showing connections and some head structures (between the pairs of pads). The platter would fit between the two heads depicted. The two large pins on the right are not part of the drive but are included for scale; the distance between their centres is 2.54 mm.*

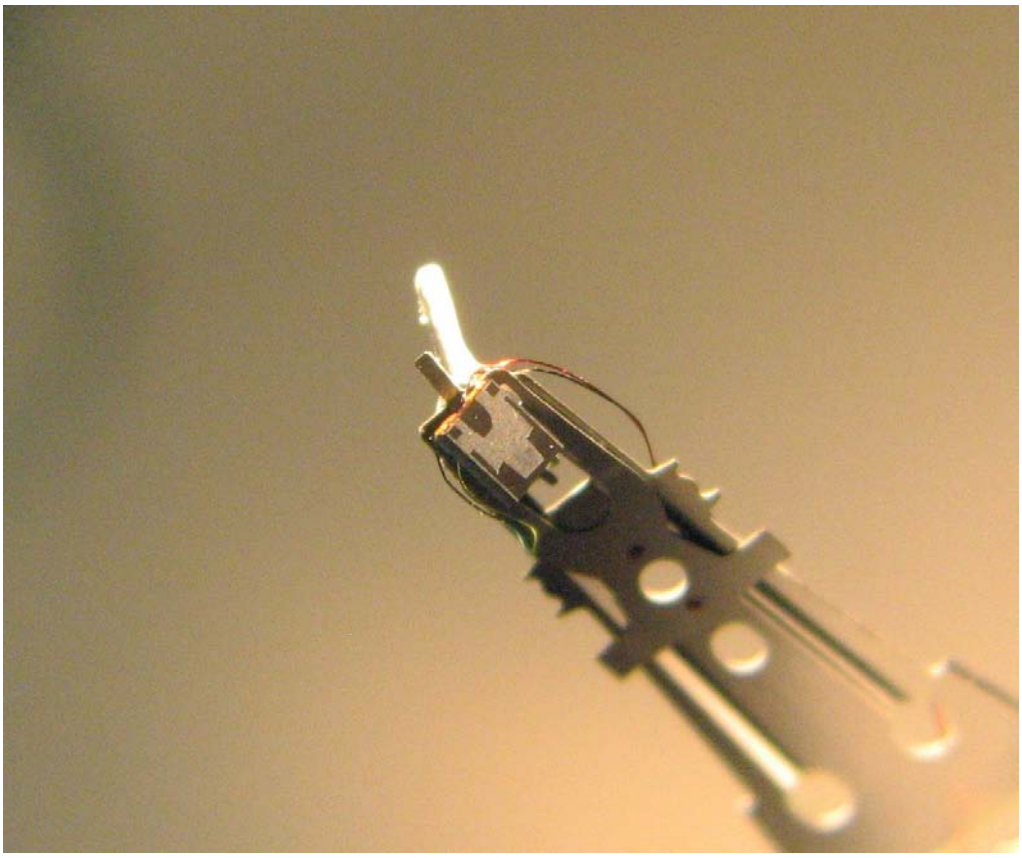


Figure 4.27: *Head slider from IBM DTLA-307045 drive. The aluminium aerodynamic 'prong' on leading edge of the head arm is fragile and has been bent upwards. The sensing area is the upper edge of the black triangle.*

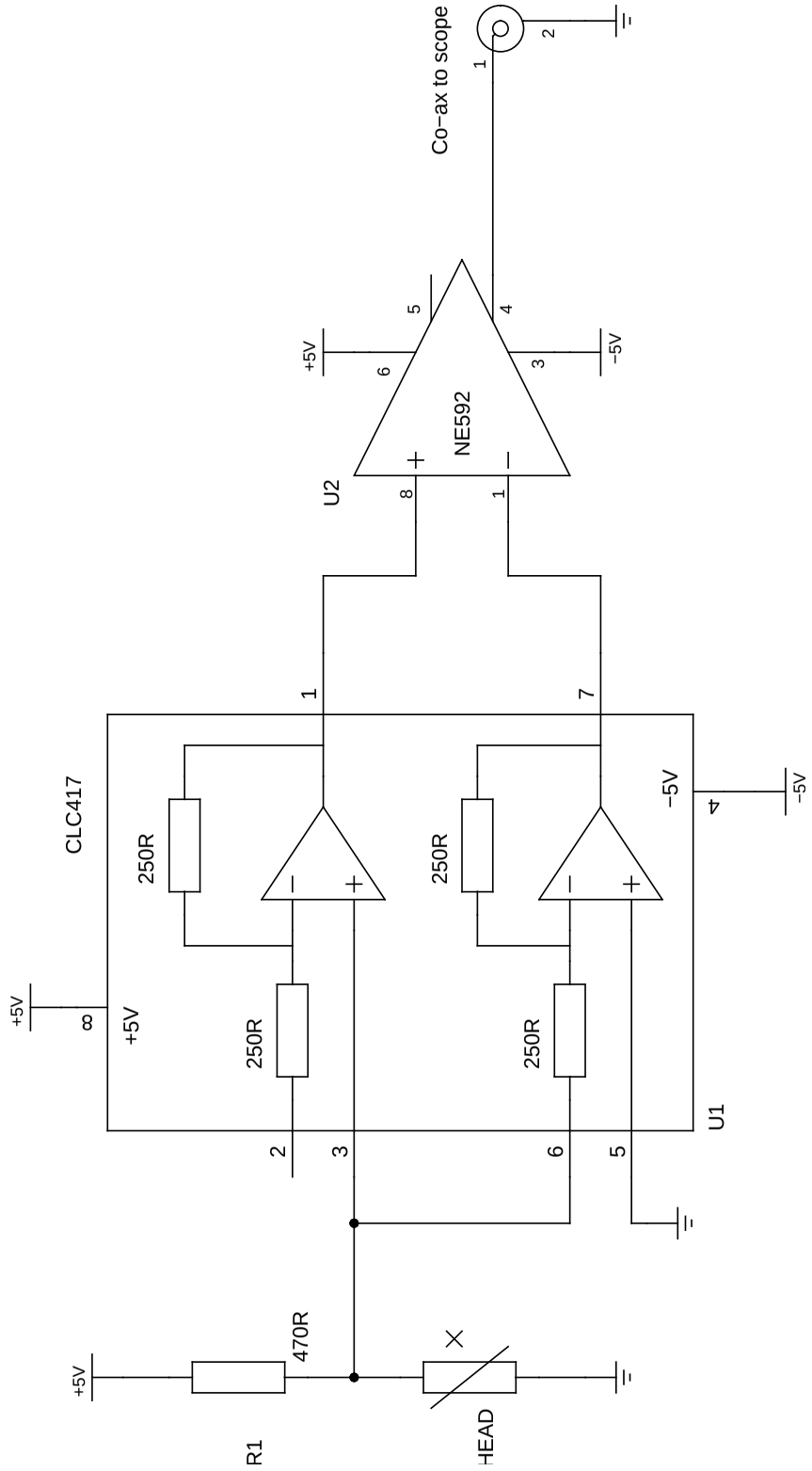


Figure 4.28: *Test circuit for IBM Deskstar GMR head*

There was only noise emanating from the output when placed near the LH77790B test chip. There was no correlation between the output and the position of the head, nor anything resembling a signal related to the processing being performed.

A high-pass filter was added, using an additional series 2.2 µF capacitor and a 120 kΩ resistor to ground between the CLC417 and NE592. Still no meaningful signal could be detected.

I also reasoned that, despite the GMR read head, most likely the write head was still inductive which might be used to sense magnetic fields. I used the same circuit with the bias current resistor removed, and tried both heads separately in this configuration. Nothing but noise was identifiable. It may be that my assumption was incorrect, that the write head is not sensitive enough, or that there is more than a simple inductive coil between the two terminals.

Figure 4.6 on page 85 suggests that GMR is only sensitive down to 10^{-1} Gauss, or 7.9Am^{-1}. The magnetic field of a printed circuit board track of width w with the sensor placed directly above the track can be approximated by (Kraus 1991, p. 240):

$$H = \frac{I_x}{2w} \tag{4.5}$$

Considering an on-chip power wire of perhaps 100 µm, we find that a GMR sensor can resolve a current of 1.6 mA. With this resolution we are unlikely to discern currents of this order through noise. A thinner track will have a stronger magnetic field for the same current, so it is likely that thin tracks do not carry large currents. Despite having obtained no data for the head under test, it seems likely that it would be suffering from these problems. Therefore it does not seem worth pursuing this avenue.

4.6 SPRINGBANK MEASUREMENTS

Simple power attacks were performed on the XOR operation of the Springbank chip. Two different programs were run on the XAP with a core loop of:

```
loadloop:
; positive trigger on IOM[0] output pin
                st      ah,@(0,x)
; load value from memory
                ld      al,@val
; negative trigger on IOM[0] output pin
                st      y,@(0,x)
                bra     loadloop
```

By modifying the data section of the program, the value loaded was set to be 0xFFFF in one program, `LoadFFFF`, and 0x0000 in the other, `Load0`.

Electromagnetic signals were averaged over 5000 sweeps with a LeCroy LC564A oscilloscope, to average out the noise power received. To minimise experimental error, each of the two programs were run twice in the order `Load0`, `LoadFFFF`, `LoadFFFF`, `Load0`. Any non-operation-dependent factors will show if the two `Load0` traces are different. I also plot one sweep of differential core power for the whole chip with no averaging. Each experiment of 4 program runs was taken over a few minutes, with longer times between experiments. 15 minutes warmup time was allowed, starting from cold.

A similar test was undertaken using the following program:

```
; begin the loop.  Set the AH register to a preset value
itloop_dotest:  ld      ah,#H'0055
                nop
                nop
                nop
                nop
                nop
; positive trigger on IOM[0] output pin
                st      al,@(0,x)
; XOR preset value in AH register with fixed 0x55
                xor     ah,#H'55
; negative trigger
                st      y,@(0,x)
; next iteration
                bra     itloop_dotest
```

The aim was to investigate the effect of an XOR with a different Hamming weight and Hamming distance. The XAP1 architecture has few registers, so it is not possible to reserve a register for the target value and it must be preloaded in each iteration (the first line of the loop). To reduce any knock-on effects from this load, it is separated from the core of the

loop by five NOPs. There are two versions of this program: `XOR55to55`, in which the first instruction loads 0x55; and `XOR00to55`, where it loads 0x00. The choice of 0x55 rather than 0xFF is less than optimal: it was chosen to only transition half the number of wires, both in single- and dual- rail logic. Subsequent tests (Section 5.4.1 on page 152) used 0xFF instead.

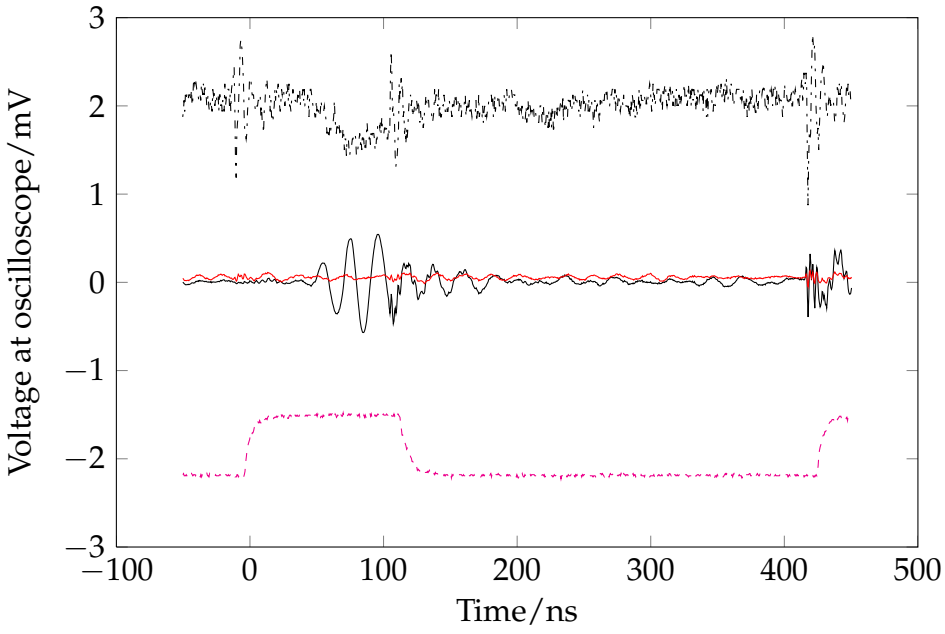
4.6.1 AC280 inductive head

The positioning of the head made a large difference in the received signal. Figure 4.29a on the following page shows a difference in EM measurements when running the programs on the synchronous XAP, with the head over this core. The `Load0-Load0` trace is very small, suggesting that there have been few external variations (movement, supply voltage changes, etc) over the run of the experiment.

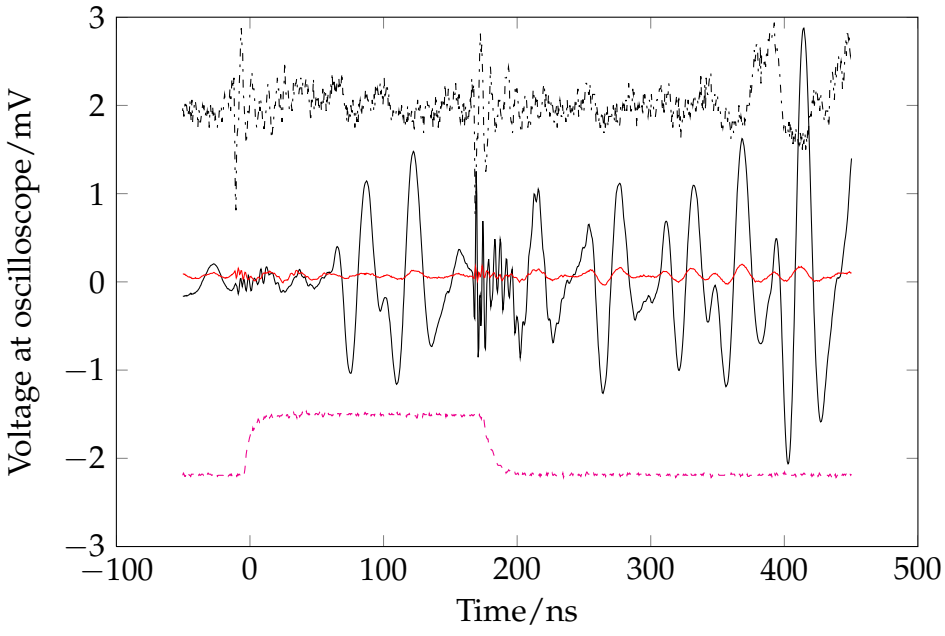
This shows a noticeable power difference blip in the middle of the trigger pulse when data is being loaded from memory. In addition, with the head over the synchronous XAP, we can see an EM difference pulse coincident with it which subsequently dies away. When over the secure XAP, this pulse may also be present, but the picture is much harder to discern.

The converse is seen when running the program on the secure XAP. Figure 4.29b on the next page shows the traces with the head over this processor. In this case, we see a huge EM difference trace starting from the point of data dependency when over the secure XAP, whilst a similar, yet much smaller, trace is seen with head over the synchronous XAP. This suggests that physical proximity is a key factor in determining the EM received.

Why is the EM difference trace for the secure XAP so much larger than that of the synchronous XAP? Both traces were taken with exactly the same gain settings, but it was difficult to ensure that the probes were the same physical distances from the chip since they were moved manually between experiments. Thus the magnitudes of the EM difference traces for each position should not be directly compared. However, we can still see that the EM difference for the secure XAP continues much longer in time than the synchronous XAP, even including differences in execution speed shown by the length of the trigger pulse. We hypothesise this is due to data-dependent timing within the asynchronous secure XAP: loading a value with a differing Hamming weight takes a different amount of time which is reflected in the EM difference trace. This then offsets in time all subsequent operations, being visible as an EM difference.



(a) Synchronous XAP



(b) Asynchronous 'Secure' XAP

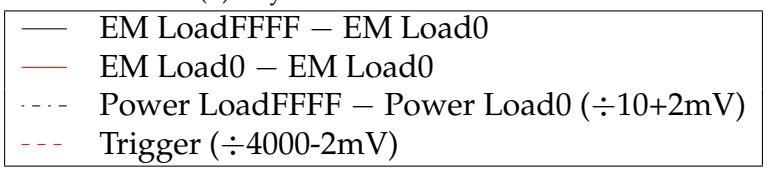


Figure 4.29: *AC280 head over Springbank, simple EM/power analysis of load instructions. The instruction of interest happens in the middle of the high period of the trigger. Both show data-dependencies, the 'secure' XAP being more long-lasting due to a data-dependent difference in timing.*

A similar effect is noticeable in the XOR test (Figure 4.30 on the following page). Both figures show noticeable EM differences, but the secure XAP's asynchronous construction means a timing shift 50 ns after the trigger.

4.6.2 HMC1002 AMR sensor

The `Load0/LoadFFFF` experiment was also performed with the HMC1002 sensor. Accurate positioning is difficult to achieve visually when the target is obscured by its large SOIC package and PCB, so the device was positioned roughly over the chip.

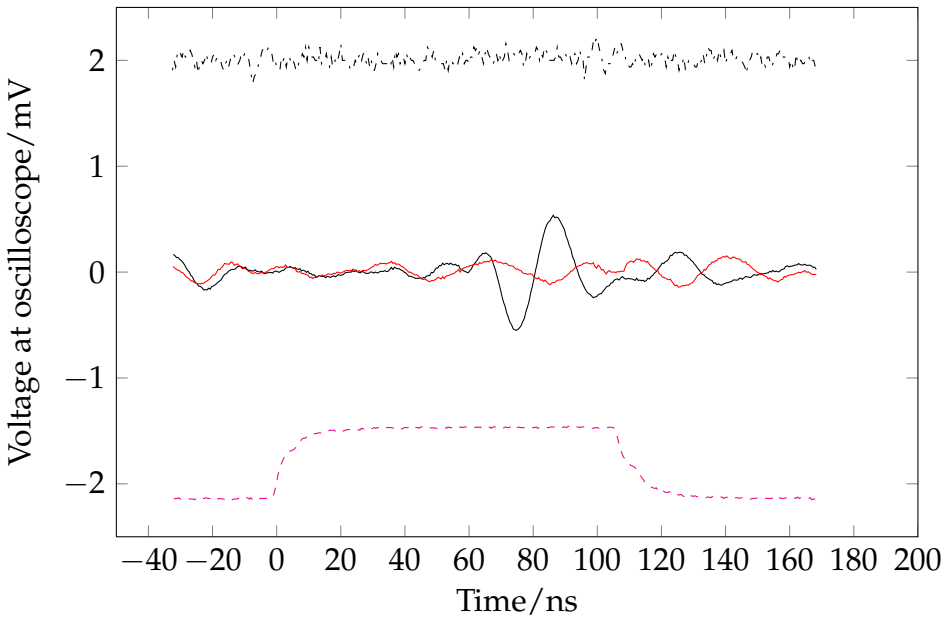
The HMC1002 datasheet does not give the distance between the two dice, which separately give X and Y magnetic fields, so relative positioning between each axis cannot be accurately judged without depackaging the device. Despite this, the two axis sensors seem to give roughly similar results, suggesting that the sensors are not very localised (Figure 4.31 on page 117). Whilst the trace for the secure XAP is larger than that of the synchronous XAP, this may be related to positioning. As it is difficult to target a particular area accurately, the magnitudes may not be directly comparable; there is also plenty of ringing from the I/O which makes this difficult to assess.

4.7 LOCHSIDE MEASUREMENTS

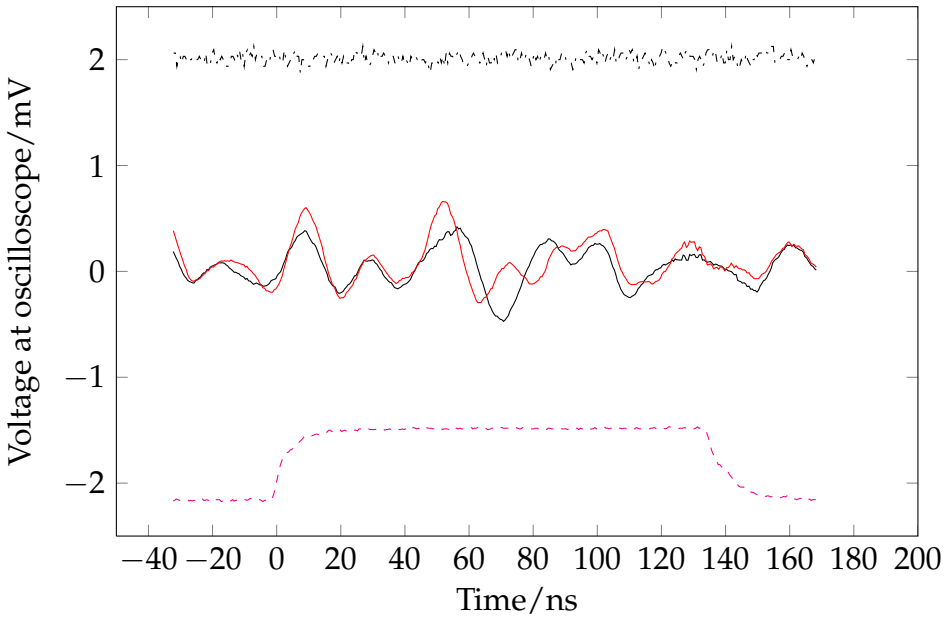
4.7.1 Lochside electromagnetic analysis

The Lochside chip was evaluated using the AC280 sensor positioned over the security test block, as close as possible to the chip surface without snagging bond wires (about 0.5 to 1 mm away from the chip).

The magnetic field was found to be highly dependent on whether a pin was being driven: pin harmonics were very visible (Figure 4.32 on page 119). All the different antennas were tried from both on-chip oscillators without the pin active, but there was no discernible difference in the magnetic field measurements. This suggests that it is the logic element switching which generates most EM emissions, rather than the wiring acting as antennas. Although the on-chip wiring was not tuned to the frequencies of the on-chip oscillators, the wiring is not an optimal antenna in any event. The oscillators can only produce a fixed set of frequencies so they cannot be tuned to the antennas (except the PLL, which needs a reference frequency input and thus itself creates plenty of EMI). Difference in



(a) Synchronous XAP



(b) Asynchronous 'Secure' XAP

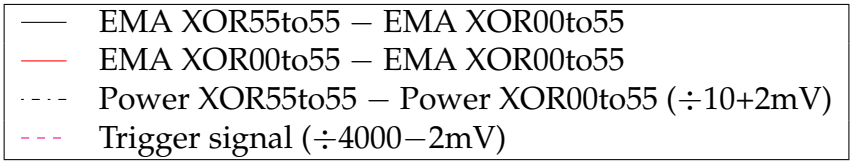
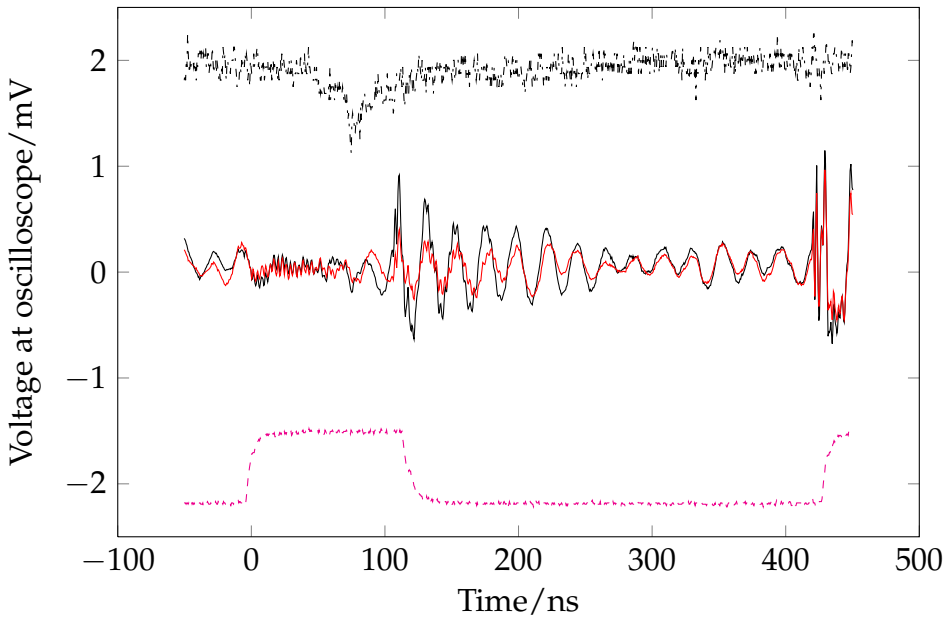
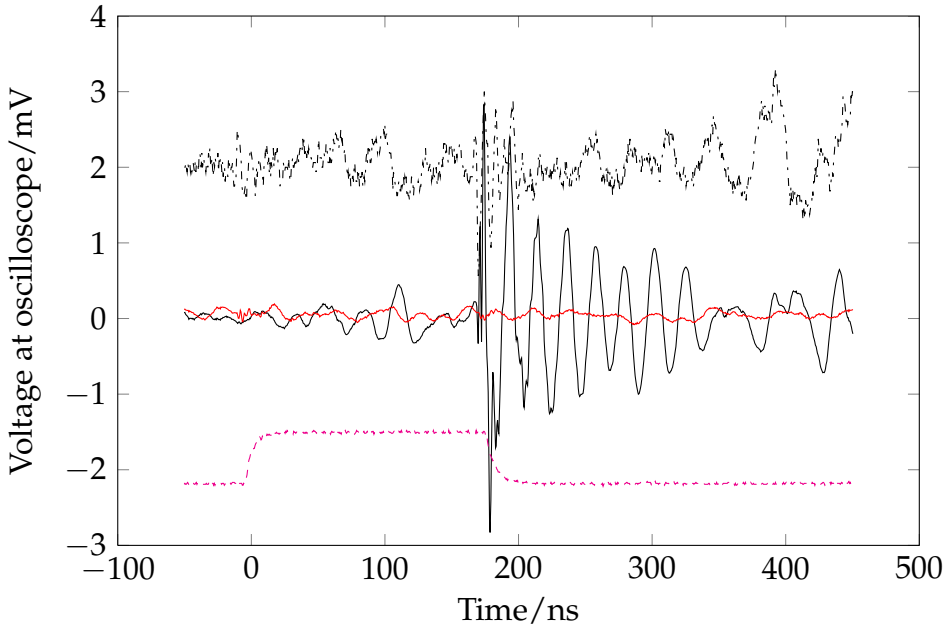


Figure 4.30: *AC280 head over Springbank, simple EM/power analysis of XOR instructions. Again a data-dependence is apparent in both cases, more pronounced on the 'secure' XAP due to timing differences.*



(a) Synchronous XAP



(b) Asynchronous 'Secure' XAP

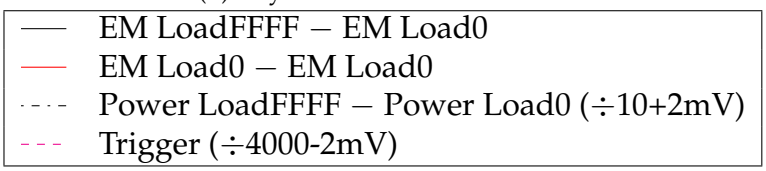


Figure 4.31: *HMC1002 over Springbank, running different loads (as Figure 4.29 on page 114). Most of the detectable differences are blotted out by the I/O trigger rather than the load instruction.*

behaviour due to the antennas could not be detected even at the highest drive strength.

The long ring oscillator was enabled and run through antenna 9, the 1.56×0.09mm Metal6 loop around the chip, which is the largest on-chip antenna we have available (and parallels the full-perimeter loop of inverters making the ring oscillator).

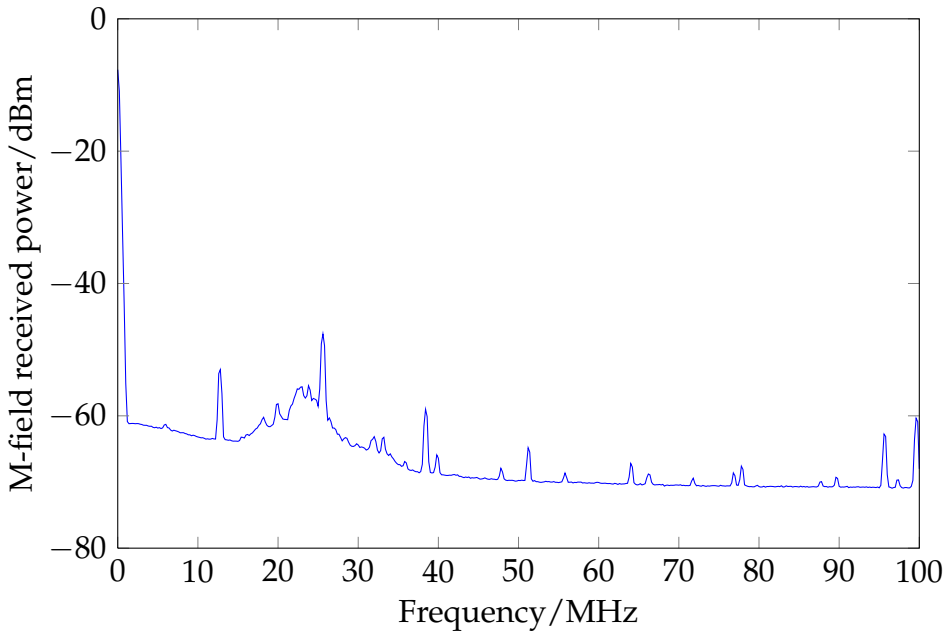
With the AC280 sensor about 0.5 mm from the die surface over the security test block, a clear set of harmonics could be seen (Figure 4.33 on page 120). The high end of the same harmonics were received on the E-field probe (Figure 4.34 on page 121).

4.7.2 Scalar network analyser from Lochside Distributed Clock Generator

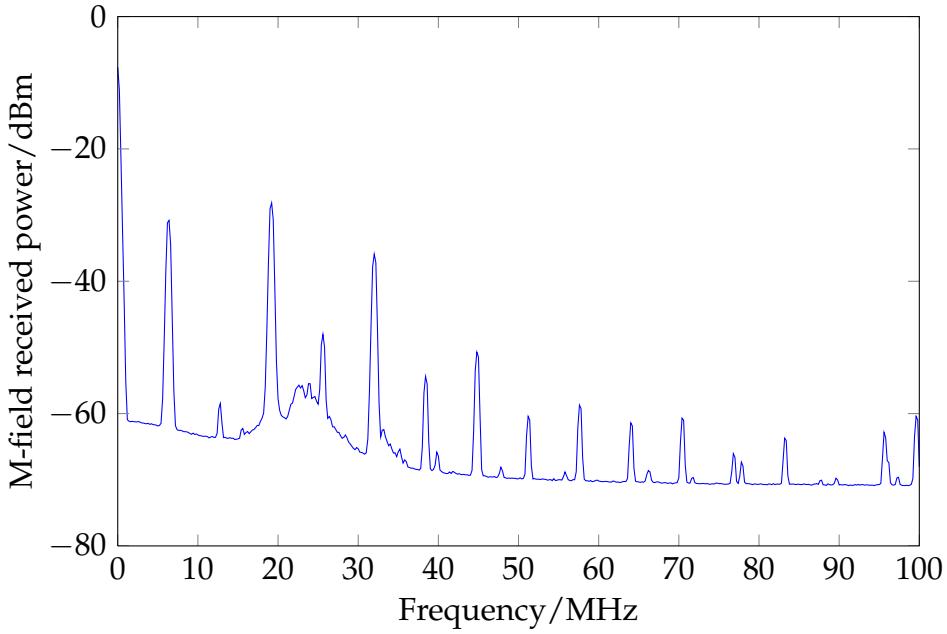
The Distributed Clock Generator (DCG) (Fairbanks and Moore 2005) is a a grid of asynchronous oscillators that lock phase and oscillate together with very low skew to form both clock generation and distribution. They are evenly spread in a mesh across the surface of the chip to replace a conventional clock H-tree. The DCG makes use of the 'Charlie effect' which spreads the tokens evenly throughout each asynchronous FIFO ring. A phase mixer at each node of the mesh is used to correct phase shifts, with the result that each node of the mesh is locked in phase and the grid functions as a single oscillator spread over the surface of the chip.

Fairbanks asserts that "the supply conductors [on the chip] provide free [EM] noise shielding", but this is questionable given other effects: they may only shield the E-field, they may re-emit, and they may couple power consumption of the chip into the spectrum of the oscillator (see Chapter 5). In theory, however, the distributed nature of the clock generator should distribute the emissions power further over the surface compared to a focused emitter. The reduced power consumption of an asynchronous circuit also reduces its emissions (van Berkel, Josephs, and Nowick 1999). A spatial spread may be of value for defence, especially if the power can be reduced below the noise floor; but the attacker can fight back by using a larger sensor to integrate the distributed emissions at the expense of integrating the noise.

We have no equivalent clock generator with which to compare (apart from the PLLs, which are difficult to control without extra emissions from the input signals), but the DCG does allow us to form a network analyser without the difficulty of the injection signal. By using the DCG as a voltage-controlled oscillator (VCO), we can measure the response of



(a) Oscillator output pin off



(b) Oscillator output pin on

Figure 4.32: *Lochside magnetic field (AC280 sensor), when driving an output pin and not driving. The harmonics from the I/O square wave are very noticeable. Long ring oscillator running (configuration 791830/793830).*

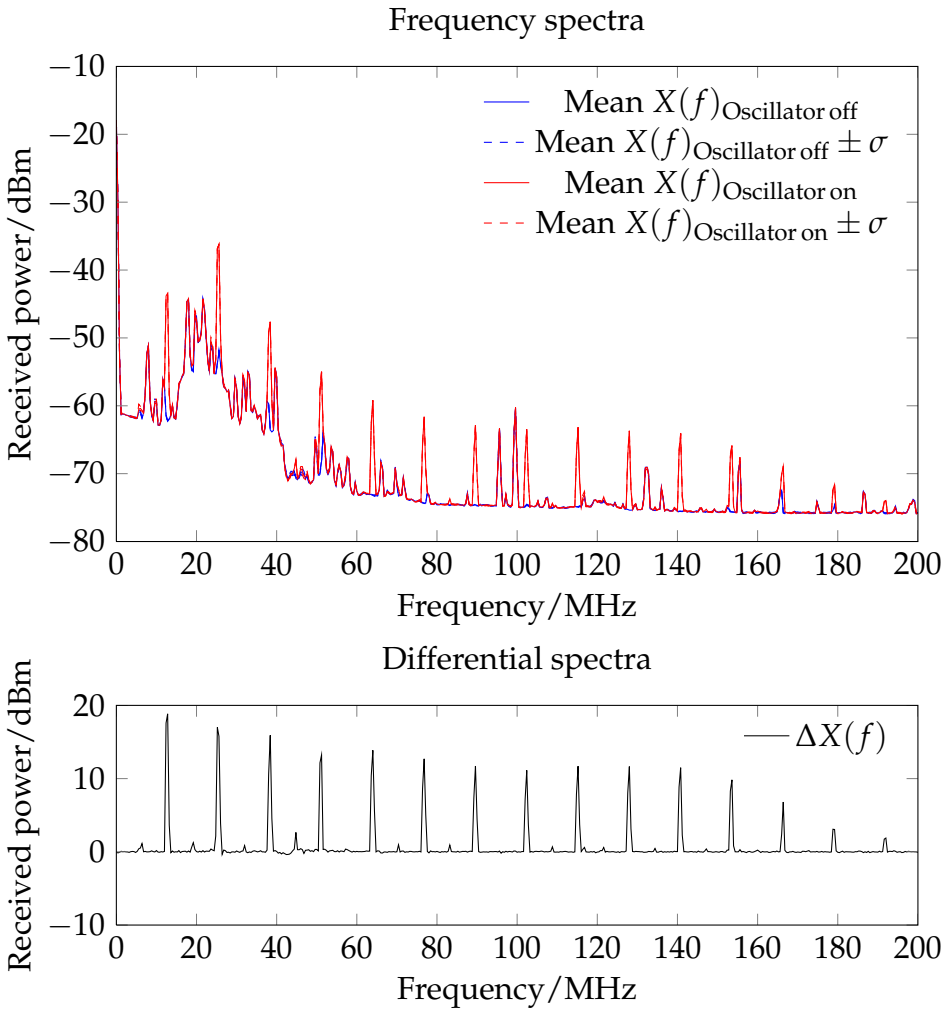


Figure 4.33: *Lochside M-field measured by AC280 over security test block: differences between long ring oscillator driving 1.56 × 0.09 mm Metal6 loop and no oscillation. No output pin active.*

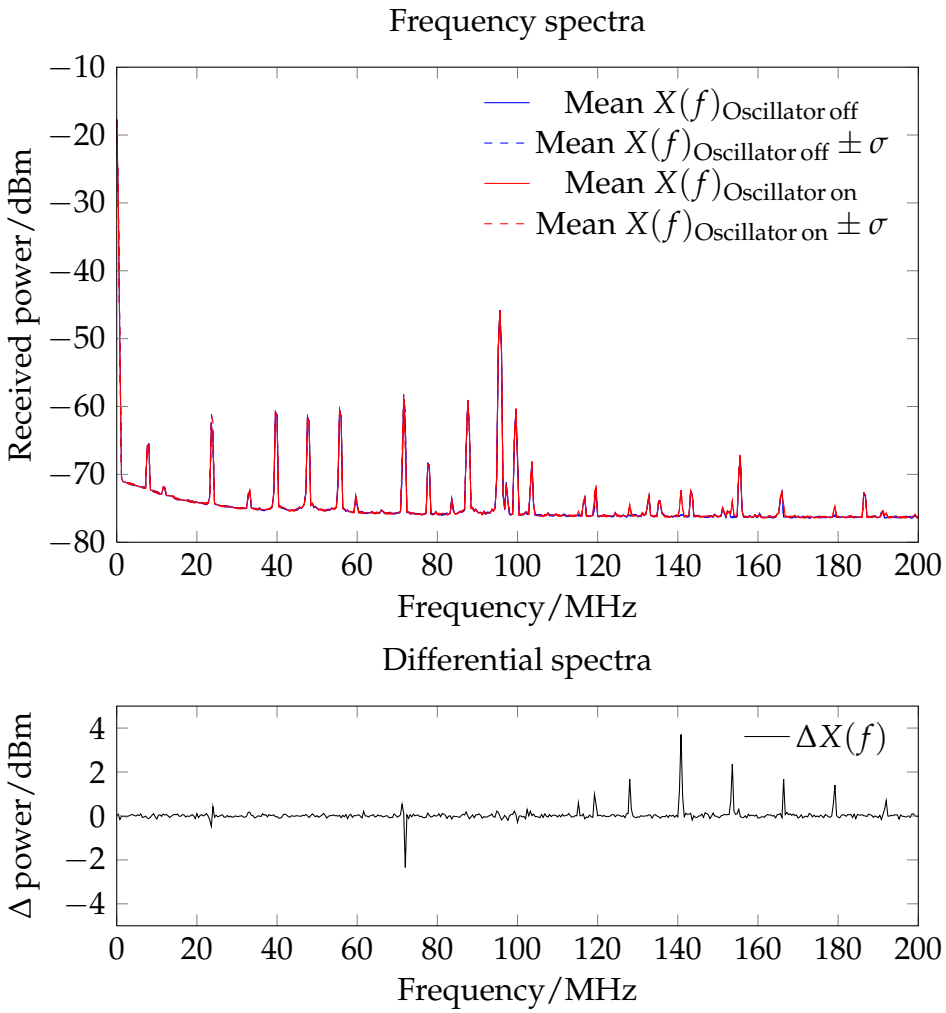


Figure 4.34: *The same as Figure 4.33 on the facing page but measuring the E-field with the brass chip cover. The E-field has a response only at higher frequencies (only relative magnitudes are comparable since the AC280 probe has a built-in amplifier).*

the wholly on-chip oscillator, the antenna, the sensor and the measurement chain without the large overprinting signal from the input pin/bond wire/pad. The DCG control voltage is a DC analogue voltage between 0.2 V and 0.9 V, to which can be added as much decoupling capacitance as necessary to prevent that pin/bond wire having any effects on the emissions spectra.

Figure 10 in Fairbanks and Moore (2005) gives the clock period for the simulated DCG network. The output of the DCG is passed through either a divide-by-8 or a divide-by-16 function before driving a pin. It is also used to clock the rest of the Lochside network-on-chip, which is disabled for these tests. In these tests the DCG was used in divide-by-16 mode to maximise the range of internal frequencies that could be measured within the lowpass filter of the pin, bond wire and driver.

The DCG control voltage was driven by a TTi QL355TP GPIB-controlled power supply in a four-terminal arrangement. One twisted pair was used to supply the current and another twisted pair to sense the voltage across the board. In this way the wiring between the power supply and the board becomes part of the voltage regulation feedback loop, and voltage drops in the wiring are compensated out. This is important because the voltage was changed in small steps of 2 mV where noise becomes significant. A 100 nF capacitor was placed across the voltage input pins on the board to attempt to reduce noise that may cause modulation of the DCG frequency (this effect is covered further in Chapter 5).

The control voltage was scanned and the frequency measured (Figure 4.35 on the next page). Also shown for comparison is the data from Figure 10 of Fairbanks and Moore (2005).

Using the DCG-generated field as the input, the frequency response of two sensors was measured. Because the DCG is a free-running oscillator, the frequency is not fixed: it depends not only on the control voltage, but also on temperature and supply voltage. Furthermore the output frequency on `DCG_OUT` contains plenty of jitter, so that a simple frequency counter will not necessarily capture the fundamental. In addition, the emissions from `DCG_OUT` will tend to swap any internal emissions. For this reason a more complex measurement strategy was devised.

First, I use the calibration curve in Figure 4.35 on the facing page to make a guess at the likely fundamental for a particular input voltage (and other conditions, such as temperature and supply voltage). I then create a window based on this frequency, plus and minus a 'target factor'. So if $f_{\text{DCG estimate}}$ =10 MHz and `targetfactor` = 20%, then f_{window} = 8 to 12 MHz. The spectrum analyser is set to sweep this window, with an appropriate resolution bandwidth chosen from a table (it has to change

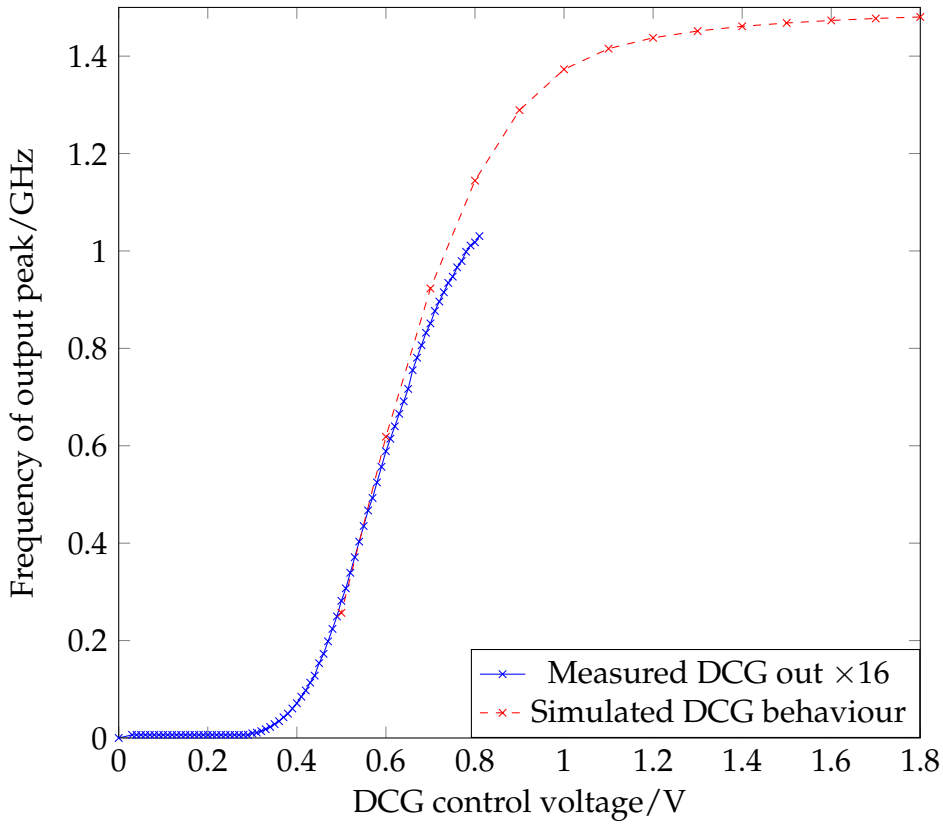


Figure 4.35: *DCG calibration curve at 1.8 V, approximately 20 °C. The DCG outputs its internal frequency divided by 16 when* `DIVCTRL`*=1. Here we plot the measured output frequency* ×16 *and the simulated results presented in Fairbanks and Moore (2005).*

since the sweep may be over four orders of magnitude, from 10^5 to 10^9 Hz).

Within this window a trace is recorded and the power of the highest peak is measured. This may be a peak unconnected with the DCG – for example, BBC Radio Cambridgeshire at 96.0 MHz – but this will be apparent since the same peak will appear on the control trace, with the DCG turned off. By comparing three traces – DCG on, output pin on; DCG on, pin off; DCG off, pin off – the responses from the DCG can be distinguished from the background signals by comparing the frequency and power of each point.

An example may be seen in Figure 4.36 on page 125. At each input voltage a spectrum in the window $f_{\text{DCG estimate}} \times (1 \pm \texttt{targetfactor})$ is captured and the frequency (a) and magnitude (b) of the highest peak re-

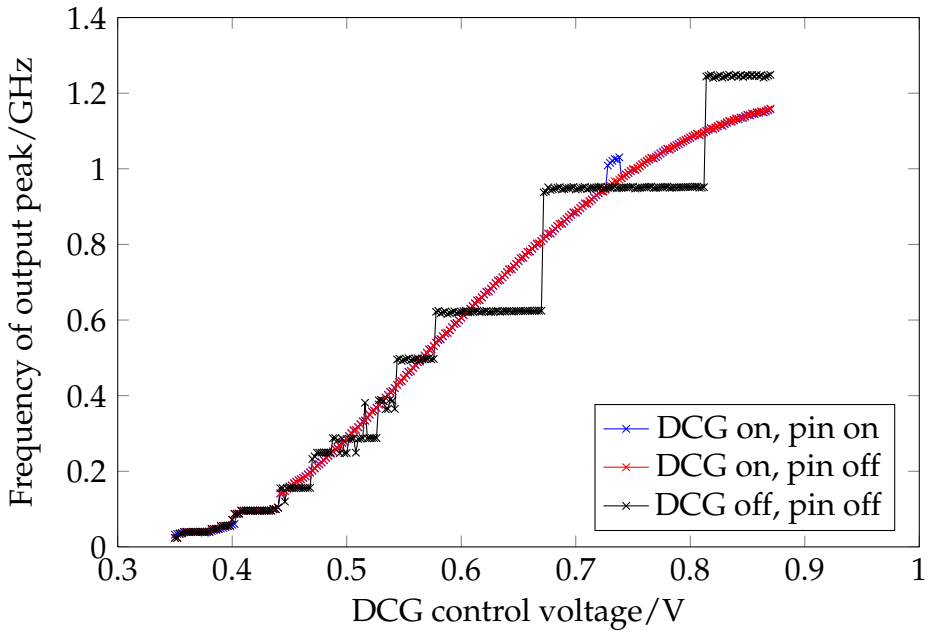
corded. To check this measurement has gathered the right data, the frequency should be a continuous curve (the red DCG on/pin on trace in (a)). If the curve is stepped, like the black trace (DCG off/pin off), it suggests the spectrum analyser is picking up a background signal which does not change frequency with changes in the DCG control voltage. This is borne out in the magnitude plot (b), when the black 'DCG off' trace has a much lower power than the red/blue traces with the DCG on. The frequency plotted against magnitude of the previous two plots are shown in (c), where the black dataset illustrates this by doubling back and forming clumps, while the red/blue datasets form a smooth curve with little doubling-back. A curve doubling back indicates a lower received frequency for a higher input voltage, which cannot correspond with the DCG output shown in Figure 4.35 on the preceding page.

The smaller `targetfactor` is, the smaller the search window; and the steps in (a) will be closer to the calibration curve even in the complete absence of any signal. In this case we must use the lower power levels and make a comparison between DCG on and DCG off traces to determine whether the response peak returned is true or illusory.

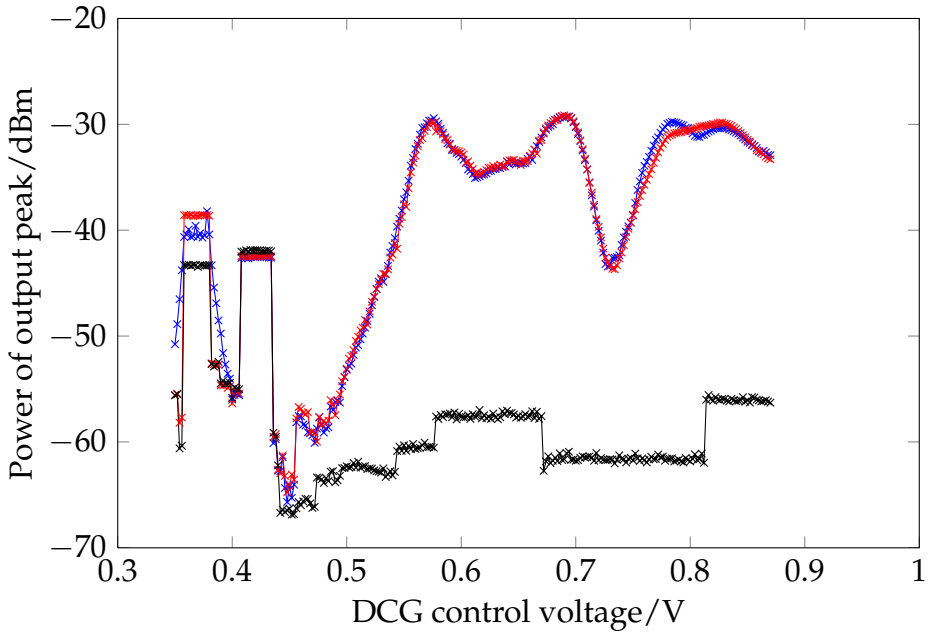
The frequency responses of the DCG with the electric field (metal lid) and magnetic field (AC280 head) sensors, while measuring the DCG fundamental f_{DCG}, and also at $f_{\mathrm{DCG}}/16$, can be seen in Figure 4.38 on page 129. When the DCG output is on, the output pin is driven by $f_{\mathrm{DCG}}/16$. Since the output is a square wave which has odd harmonics, it should not cause a large harmonic at f_{DCG} and any such harmonic is likely to come from the internal operation of the DCG. When the output is off, all signals are from the internal frequencies and the signal of interest is only f_{DCG}. There should be nothing to see at $f_{\mathrm{DCG}}/16$ unless there is leakage through the divider.

Figure 4.37a on page 128 and Figure 4.37b on page 128 show that there is no signal at $f_{\mathrm{DCG}}/16$ when the output pin is off. An upper bound on the noise floor in the sampling window is given by the black (DCG off) lines on each plot: the background harmonics plotted (which may be noise, or interferers such as FM broadcast signals) will be on average the same or greater than the noise floor would be in the absence of interference.

The magnetic sensor has most of its sensitivity between 0 and 40 MHz with a peaky response, while the electric field sensor operates up to at least 70 MHz, with a single smooth peak at 40 MHz. The electric field sensor also has a higher power, though this is due to the much larger size of the magnetic field sensor and different amplification applied to each. Magnitudes are not directly comparable, except for the fact that the noise floor of the spectrum analyser is the same in each case.

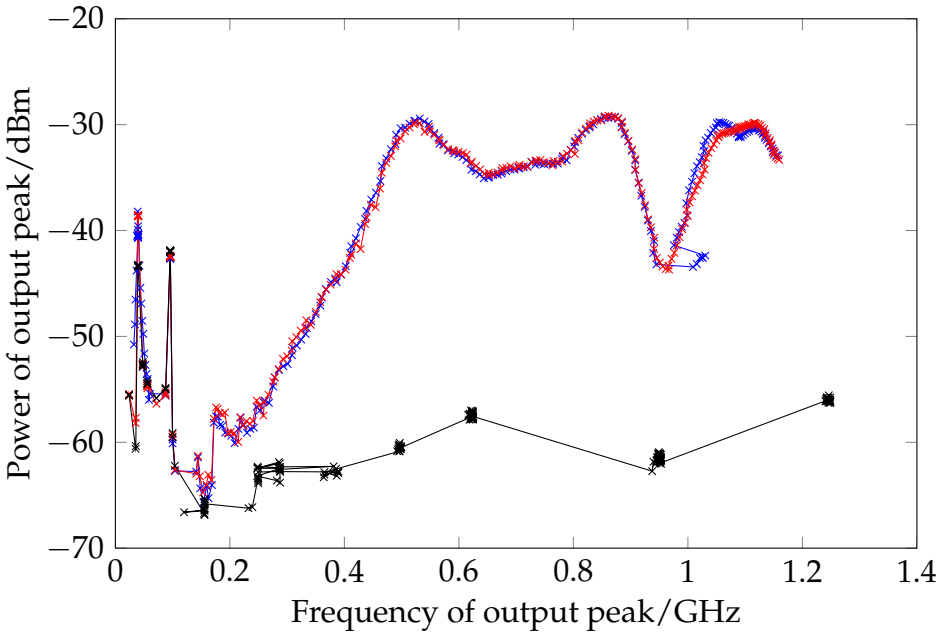


(a) Frequency



(b) Magnitude

Figure 4.36: *Construction of a frequency response curve from scans of the DCG control voltage (electric field, looking at internal DCG fundamental).*



(c) Composite frequency response with X-axis from (a) and Y-axis from (b)

Figure 4.36: *(continued).*

Looking at the fundamental f_{DCG}in Figure 4.38a on page 129, the magnetic field, there is clearly no signal to see. There are small response peaks when the DCG output is on at about 400 MHz and 750 MHz, but these are most probably artifacts of some kind as they do not look regular enough. This is not too surprising, given the low-pass filtering effect of the inductive sensor and ferrite core.

Figure 4.38b on page 129 shows much clearer electric field response curves with the 'output on' and 'output off' curves being almost identical, as would be expected. In particular, the flat response from 500 to 900 MHz is very useful, as is the large height over the noise floor. The response that cannot be distinguished from noise at $f_{DCG} < 200$ MHz is most likely because the DCG has much jitter when $f_{DCG}/16 < 10$ MHz, spreading its spectrum.

The DCG is also a distributed circuit and, as such, the metal lid integrates the electric field over the area of the whole chip and pins, while the hard drive head targets a very specific area. Section 4.7.4 on the facing page explores this in more detail.

It is evident therefore that the electric field sensor (the metal chip lid) is superior. Obviously it is much larger in area than the AC280 head but,

ignoring the magnitude scale, the wideband and flat response is most useful for EMA recordings. Both sensors are useful for different classes of measurements.

4.7.3 Power supply frequency injection results

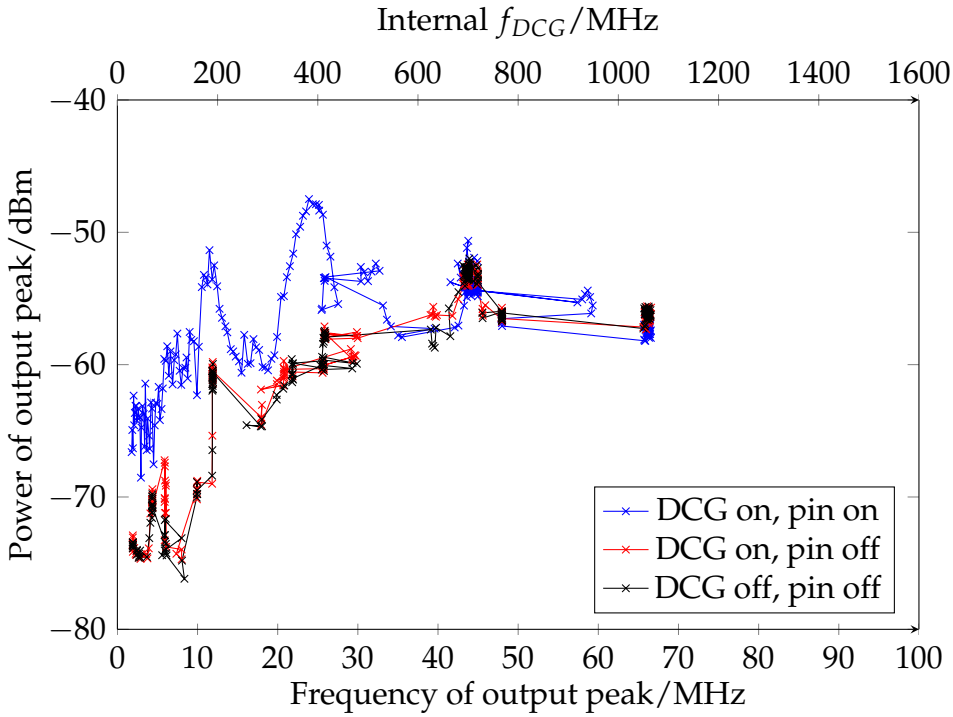
I also tried injecting a frequency into the Lochside security test block's power supply via a coupler (see Section 5.2.1 on page 143), and noting how much was visible in the electric field. The output from the vector network analyser may be seen in Figure 4.39 on page 130. The response is very strong, especially in the 0 to 1 GHz range. I developed this result into frequency injection attacks, which may be seen in Chapter 5.

4.7.4 3D scanning

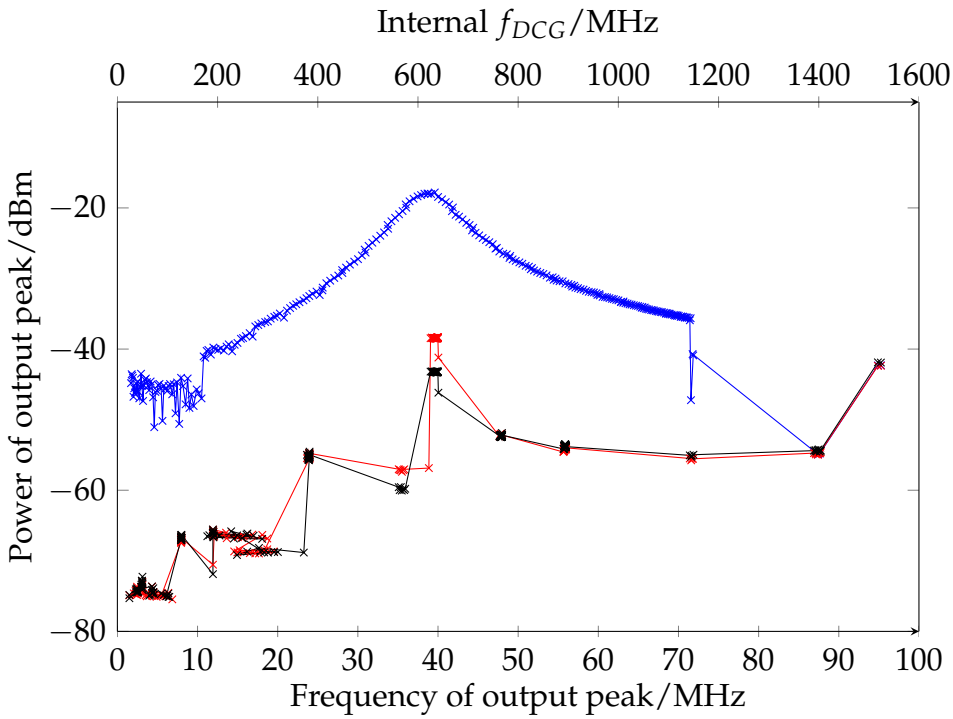
The spatial resolution of the EM sensors is potentially of interest to an attacker. For example, a computation vulnerable to power analysis in one area of the chip may be counterbalanced by another computation elsewhere. Thus the power consumption is not data-dependent, but the location of the consumption is. This may be detected by measuring EM in a specific location, but not over the whole device. Alternatively it may be of interest to pinpoint particular locations at which the key is stored, or cryptographic rounds performed, which may then be targeted with laser fault injection. Therefore it is useful to be able to listen to specific areas of the chip.

Accordingly I investigated the spatial field of the Lochside chip, using the AC280 sensor. The aim was to determine the spatial resolution of the sensor, and how far the field could be localised.

The sensor board was mounted on a metal bracket. The bracket was attached to three computer-controlled stages (together the Newport PM500-XYZ) which provides X-Y-Z control with the characteristics shown in Table 4.3 on page 130. A small manual stage with vernier adjustment was added for hand calibration. The arrangement may be seen in Figure 4.40 on page 131. The resolutions given for the motorised stages are based on the minimum encoder increment, but these are typically non-achievable due to frictional effects: for example, hysteresis and a limit of the minimum incremental movement (Newport Corporation undated). Full data for the stages was not available, but, as can be seen, the resolutions given are a thousandfold smaller than required in this application, so they are unlikely to be a problem.

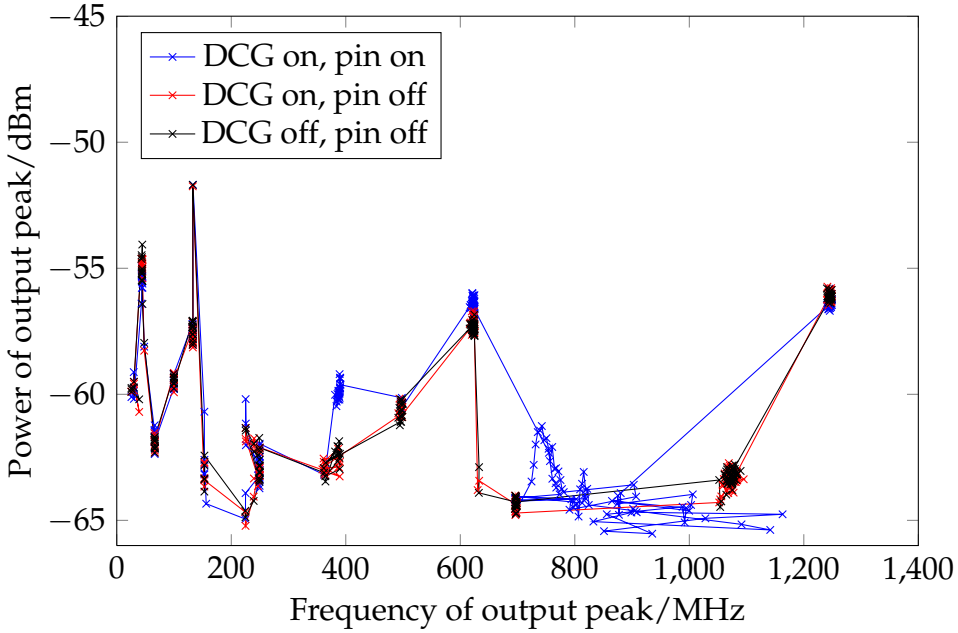


(a) Magnetic field (AC280 sensor)

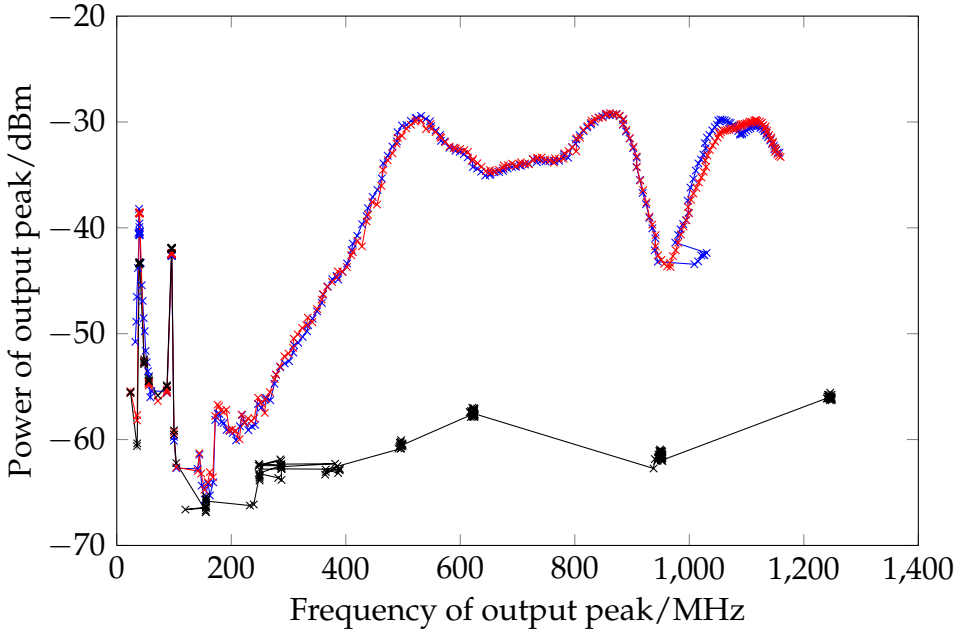


(b) Electric field (metal lid sensor)

Figure 4.37: *Frequency response curves of DCG to electric and magnetic field sensors, measuring DCG output pin frequency ($f_{DCG}/16$), with internal frequency on upper axis*



(a) Magnetic field (AC280 sensor)



(b) Electric field (metal lid sensor)

Figure 4.38: *Frequency response curves of DCG to electric and magnetic field sensors, looking at DCG internal frequency* f_{DCG}

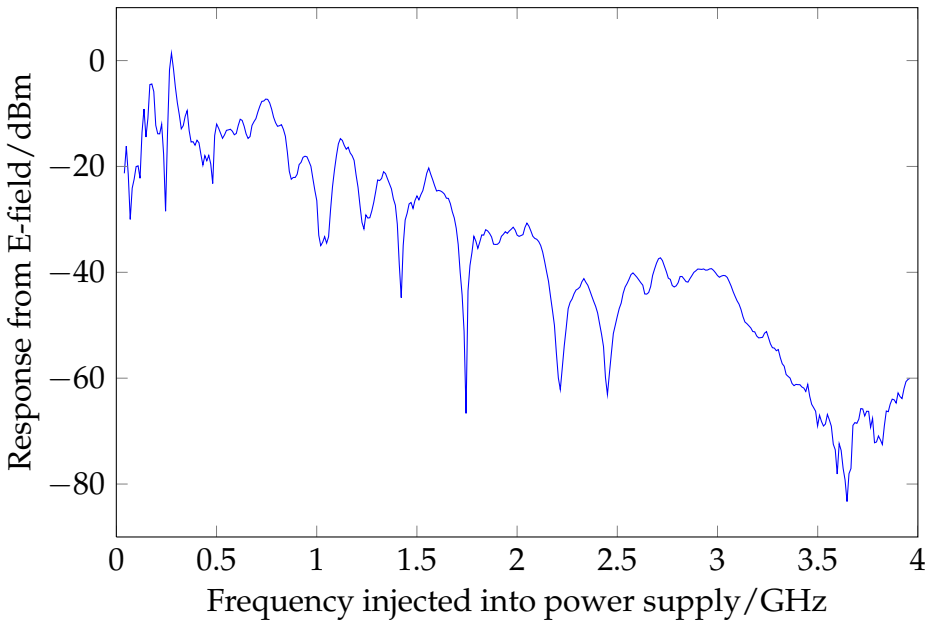


Figure 4.39: *Gain plot from vector network analyser injecting frequency into Lochside power supply and measuring response from electric field, shown relative to the input signal of magnitude −10 dBm. The coupler in Figure 5.1a on page 144 was used. The gain is very good, especially in the below 1 GHz.*

Axis	X	Y	Z	XYZ
Type	Motorised	Motorised	Motorised	Manual
Manufacturer	Newport	Newport	Newport	Quarter
Model	Mini-1 / PM500-1L	PM500-1L.25	PM500-1V.50	XYZ500MIM
Travel	25 mm	25 mm	25 mm	12.5 mm/axis
Resolution	25 nm	25 nm	50 nm	10 µm

Table 4.3: *Specifications of XYZ stages used*

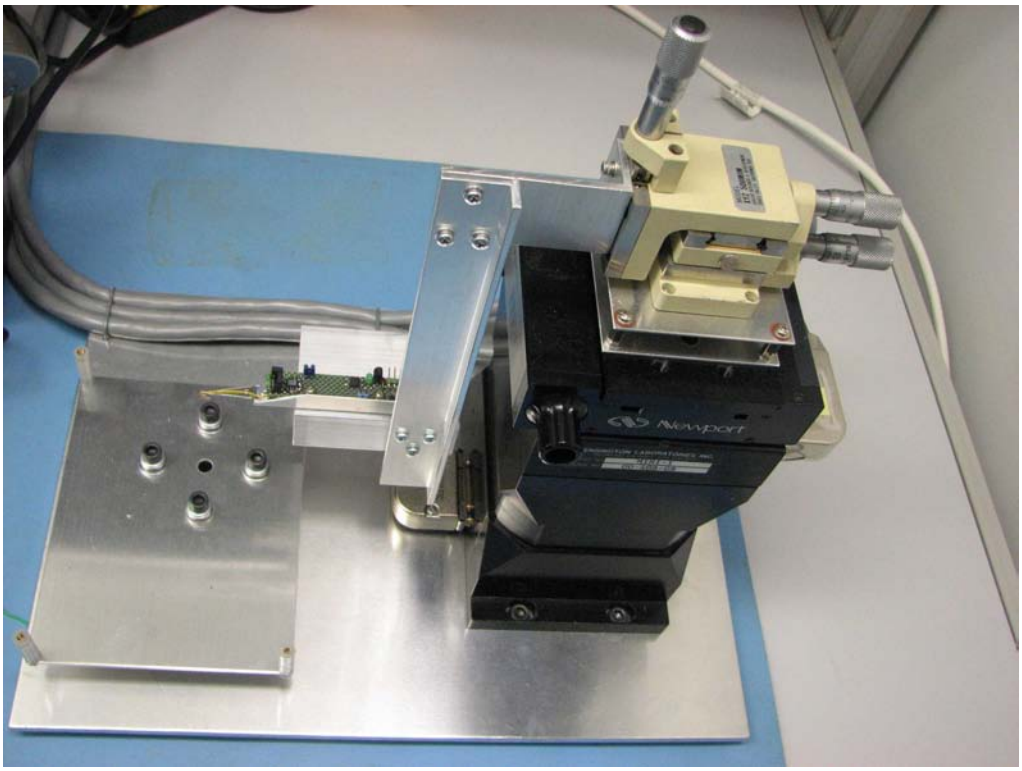


Figure 4.40: *Stack of three X/Y/Z motorised stages combined with three X/Y/Z manual vernier stages. On the arm is mounted the revised AC280 sensor and amplifier from Figure 4.15 on page 93. The board under test is mounted on the plate to the left.*

First the field of an unidentified off-the-shelf inductor was scanned with the AC280 head, using the ASK method described in Section 4.4.3.1 on page 94. The TDS7254B oscilloscope measured the root-mean-square (RMS) voltage of the received 1 MHz ASK carrier signal by subtracting the RMS of the background noise signal when not keyed (no carrier) from the RMS of the signal when the carrier was active. This measurement removes some of the contribution from the background noise (at the expense of a uniform underestimation of the RMS of the carrier).

Figure 4.41 on the facing page shows the coil and the results of 2D scanning. The same coil was also 3D scanned. A $49 \times 49 \times 28$ point volume was taken and a visualisation of the isosurfaces may be seen in Figure 4.42 on page 134.

The difficulty with using the ASK method on a chip is the recurring problem that the wiring will emit much more than the chip. So, just as for the DCG measurement, a frequency-selective method was used. For this we fix the sensor over the chip area we are interested in. We then search for frequencies which vary depending on some property of the circuit we want to measure (e.g. clock divider steps or DCG control voltage). When the chip block is turned off, the peak must disappear. When we have found one of these, we adjust the input conditions to provide maximal received power in a clear area of the spectrum. This means the peak is as large as possible, with no nearby fixed-frequency peaks from background interference. We adjust the spectrum analyser so this peak is in the centre of the window, with no other large peaks onscreen. We can then measure the height of this peak in locations across the chip surface as a proxy for electromagnetic leakage of that chip block.

To reduce measurement variation, the spectrum analyser is programmed to measure a sequence of four traces and compute their mean. It then reads off the height of the tallest peak within the window, and returns that as a single data point, in addition to the frequency of the peak. As with the DCG measurements, the frequency indicates whether we measure the correct peak or some other erroneous peak that is stronger. We can then perform a 2D or 3D scan of the chip surface. For 2D scans we position the head as close as it will go to the surface without snagging bond wires, at a distance of about 0.5 mm.

Various scans were performed: with the DCG running (Figure 4.43 on page 135), with the long ring oscillator running (Figure 4.44 on page 136) and both short and long ring oscillators running (Figure 4.45 on page 137). It proved to be difficult to see any signals coming from the Lochside core as opposed to the pins.

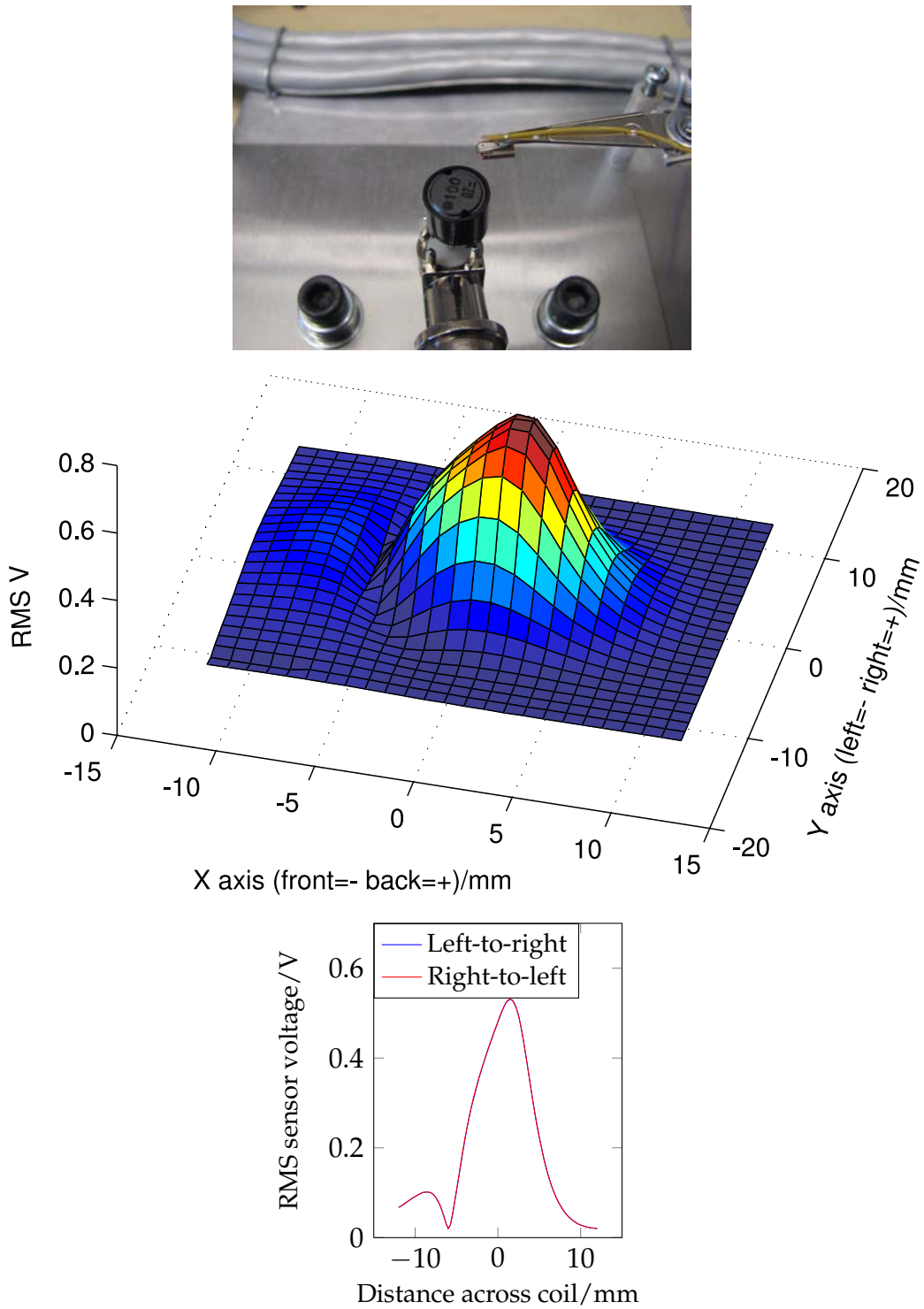


Figure 4.41: *Two-dimensional scan of inductor coil for calibration purposes, at 1 MHz on 10 kHz ASK. Note in particular how the coil has been unevenly wound, showing a dip in the field strength at the left edge. The bottom plot is a cross section scanned in both directions: the two traces overlap almost identically, showing the repeatability of the scanning and ASK measurement method.*

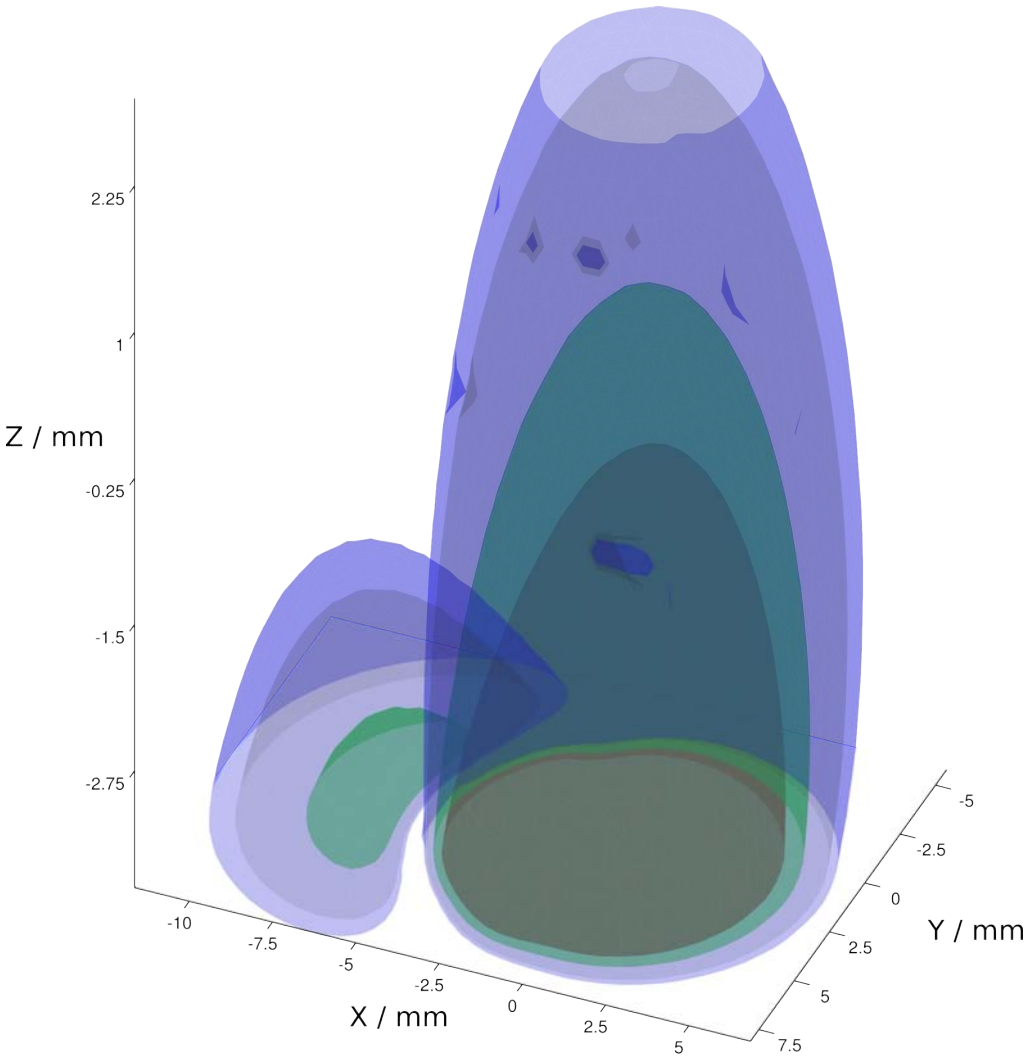


Figure 4.42: *Three-dimensional* $49 \times 49 \times 28$ *point volume scan of magnetic field of test inductor in Figure 4.41 on the previous page. Four isosurfaces are plotted:* $V_{RMS} = 0.1$ *(light blue), 0.12 (purple), 0.2 (green), and 0.3 (dark grey) volts. The extra sidelobe is clear.*

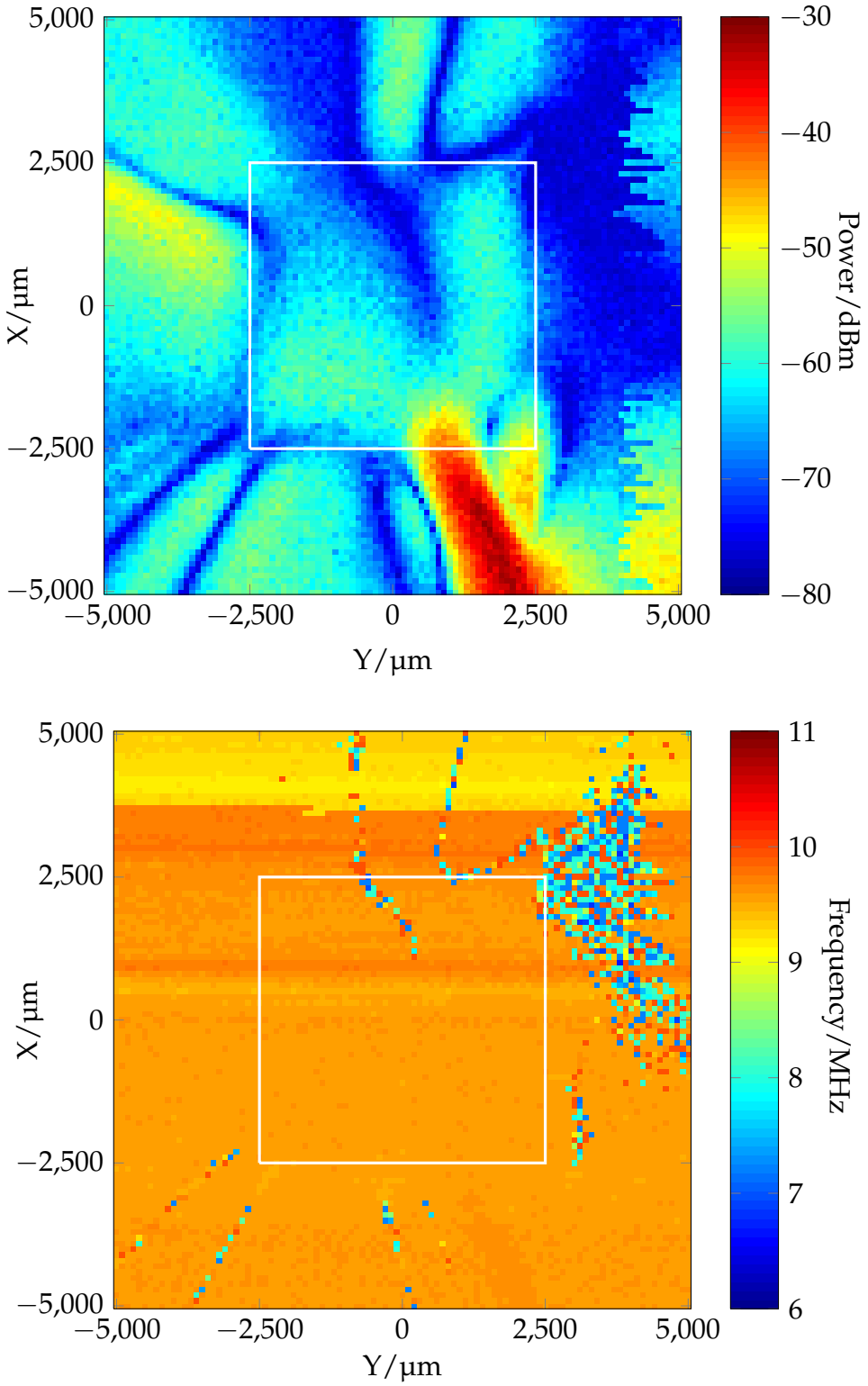


Figure 4.43: *2D scan of Lochside chip surface with DCG running with I/O pin on, V_{DCG} = 0.4 V. The box indicates the location of the 5 mm square die. The upper plot measures power of the 8.3 MHz fundamental output. The emissions are dominated by the I/O bond wire. The lower plot shows the measured frequency, indicating that the spectrum analyser measures the (low) height of another peak if the fundamental cannot be detected.*

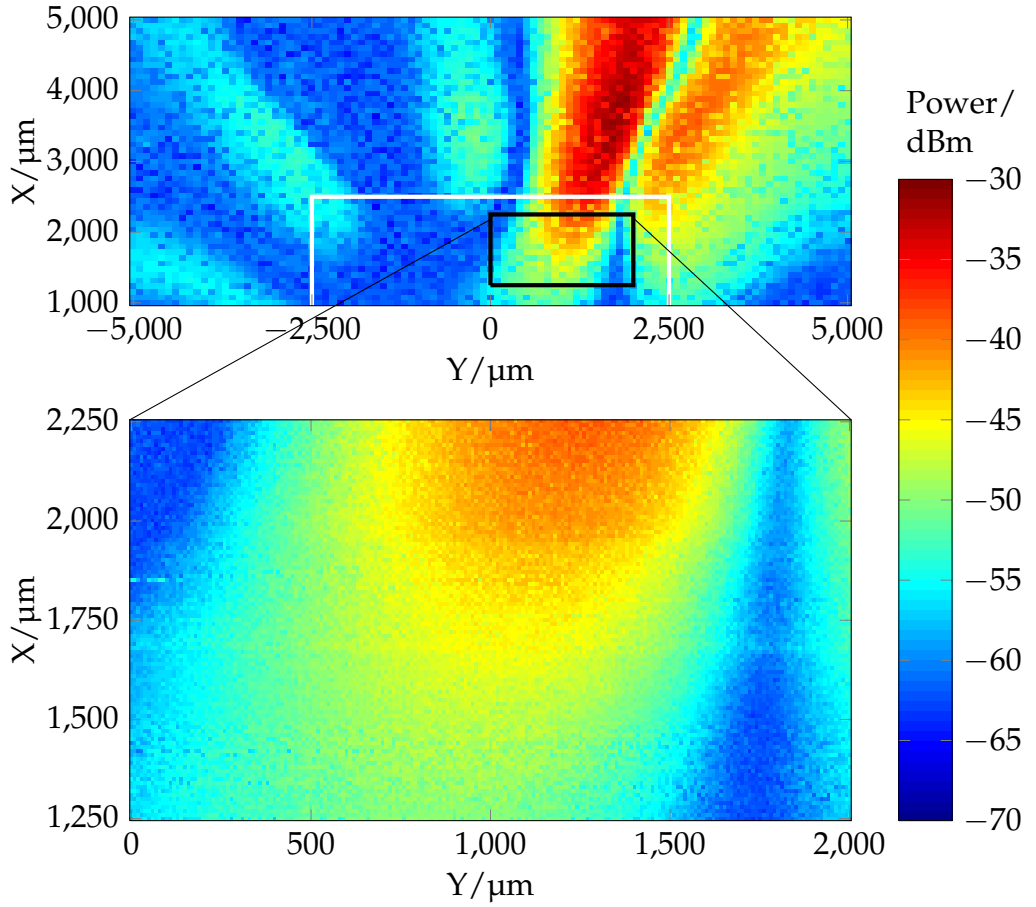


Figure 4.44: *2D scan of Lochside chip surface over the security oscillator block with long ring oscillator running (configuration 791030), measuring height of 12.0 MHz harmonic. The white box indicates the die area. With no explicit I/O, it appears that the oscillator's power consumption is illuminating the power wire. The black zoomed area is over the oscillator and pad area, with little to be seen at higher 10 µm resolution.*

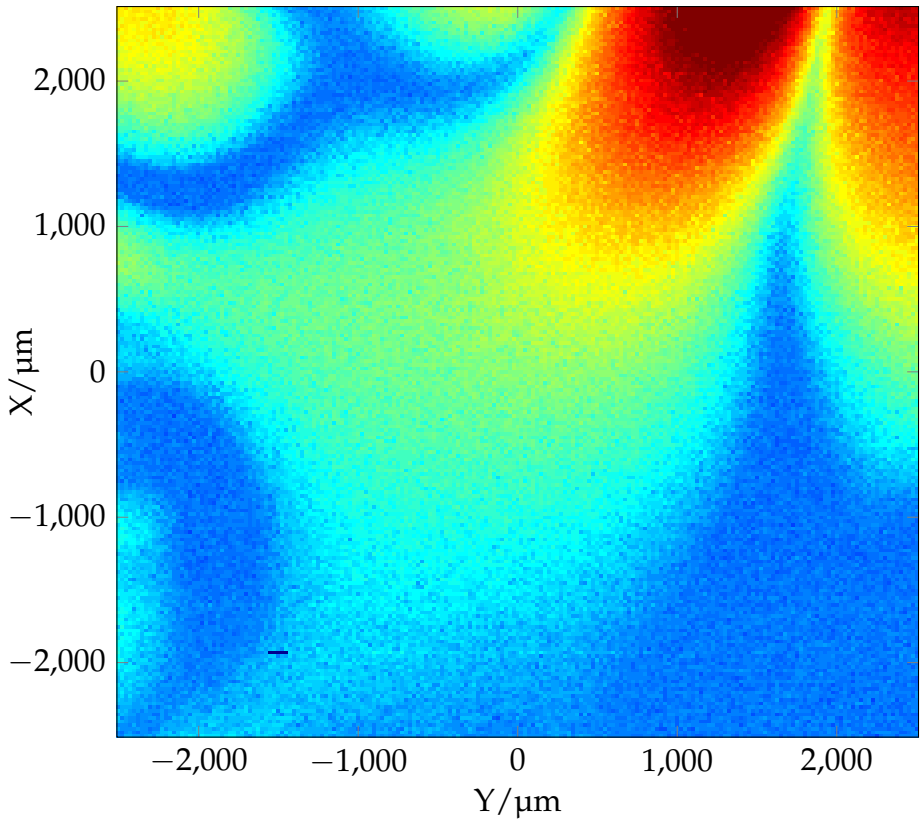


Figure 4.45: *2D scan of Lochside die surface at 25 µm resolution. Short ring oscillator is running and its output is divided by 32 then drives the long ring's chain of inverters (configuration 7D8425). Plot displays magnitude of 23.0 MHz harmonic over the die surface. Again there is little to be seen that is not an emission from the power pin.*

4.8 CONCLUSION

Much of the pre-existing work on EMA has focused on off-the-shelf probes. While these are more repeatable, they are typically not available in small sizes designed to target a small part of an IC. Quisquater and Samyde (2001) mention use of a 40µm coil (an RF inductor) for mapping a chip's emissions, but do not demonstrate its use in cryptography. Emissions mapping can use other methods (such as bus exercising or eddy current induction) to substantially increase the gain, but these are not readily applicable to a real attack. They use a 2 cm coil to perform an EMA attack, but give minimal detail of their experimental setup which does not aid in building on their work.

Gandolfi, Mourtel, and Olivier (2001) said that they had success with hard disc heads, integrated inductors and magnetic loops, but found wire coils the most useful. They do not provide any more details of these experiments, or name the sensors they tested. Again it is difficult to build on or compare their work.

I have investigated examples of inductive, anisotropic magnetoresistive (AMR) and giant magnetoresistive (GMR) magnetic sensors, and two electric field sensors. I concluded that GMR heads were not sensitive enough, and the AMR sensor under test was bulky and not ideally suited to a test environment. There also remain significant questions about the efficacy of my tests given the need for substantial guesswork and reverse engineering of manufactured products. The metal lid capacitor electric field sensor and the AC280 hard drive head were found to have the best performance within the restricted set of sensors I was able to measure. This agrees with the results of Agrawal, Archambeault, Rao, and Rohatgi (2002b) who found foil electric field sensors most useful.

On Springbank, inductive sensors had the best performance and were used successfully to distinguish memory loads of different Hamming weights. This distinction was more pronounced on the asynchronous test processor than on the synchronous processor, possibly due to data-dependent timing. The metal lid sensor was fabricated after these tests when the original equipment was not longer available.

On Lochside, the metal lid sensor had a better response at high frequencies (up to 1 GHz) compared with the AC280 head. The AC280, however, gave much better performance at low frequencies (below 100 MHz).

The spatial plots of Lochside magnetic field implied that most of the emissions were coming from the power bonding wires, rather than from the chip itself. This may be the result of testing a chip that had relatively little switching activity and with an oscillator designed for low EMI; but

it bears out the idea that EMA, in this instance at least, has little extra to contribute compared with pure power analysis.

Sauvage, Guilley, and Mathieu (2009), published after I concluded the cartography experiments in this chapter, found that commercial field probes were more convenient than self-designed ones. They found that electric probes had no positive results, which is interesting given the success both I and Agrawal, Archambeault, Rao, and Rohatgi (2002b) had. One difference is that both our sensors were whole-chip sensors, while the E-field probe used by Sauvage et al. were smaller and thus less sensitive. They opted for a 0.5 mm magnetic field probe, which is similar to that used by Gandolfi, Mourtel, and Olivier (2001). They perform cartography in pixel squares of 400µm per side, which is significantly coarser resolution than I used.

Their attack is perhaps aided by the choice of target as an FPGA, which will have more harmonics than my simple oscillators. They achieve broadly similar results by detecting what they believe to be the power and ground networks, the clock trees, pin amplifiers, wire bonds and other 'infrastructure' features. Their choice of an also FPGA allowed them to explore further the effect using different designs of cryptographic modules and perform real cryptographic attacks. This would have been another avenue I would have pursued had more time been available.

Despite their limitations, the experiments in this chapter provided a series of sensors that I then used in subsequent active attacks.

CHAPTER 5
THE RE-EMISSION ATTACK

5.1 PRINCIPLES

Agrawal, Archambeault, Rao, and Rohatgi (2002a), one of the first systematic EMA papers, characterise emissions into the following categories:

1. Direct Emanations, resulting from intentional current flows; and
2. Unintentional Emanations, caused by coupling between elements in close physical proximity, notably:
   (a) Amplitude Modulation: a non-linear coupling (such as a diode junction) between two signals causes one signal to be amplitude modulated onto the other. A receiver tuned to the carrier frequency can demodulate the data.
   (b) Angle Modulation: circuits may be coupled in by phase or frequency modulation. For example, a ring oscillator or an asynchronous circuit has a natural frequency of operation which depends on its power supply voltage. The data-dependent power consumption of another circuit may frequency modulate the free-running frequency of the oscillator. A receiver with a suitable bandpass filter may be able to extract the signal, and measure its original transmitted strength.

Agrawal goes on to demonstrate several ways in which the direct and unintentional emissions have been measured from existing smartcards. The 'unintentional' modulation-based leakage is worth studying. A smartcard may leak valuable information in the EM spectrum by modulating some frequencies of its operation. The difficulty for the attacker is the practicality of receiving it. It may be of a low frequency, requiring infeasibly large reception antennas. It may be weak or impeded by shielding.

So let us consider if an active attacker can assist in this process. What if they inject unwanted frequencies into the device, in the hope that the device will modulate them with important information? The modulation frequency would be carefully chosen for optimal propagation and detection at the attacker's antenna. The modulated frequencies could then be detected, demodulated and decoded to reveal the secret data.

This attack is something of a hybrid of the techniques described in Section 2.11.2 on page 38: a TEAPOT (malicious EM) attack aimed to make use of NONSTOP/HIJACK (unwanted cross-modulation) phenomena.

5.1.1 Coupling modes

Electromagnetic coupling is defined as "the coupling of an electric, magnetic or electromagnetic field from one conductor into another" (Weston 2001, p. 121). Coupling modes can be conducted or radiated. The conducted mode involves an electrical connection between the conductors, while the radiated mode implies a near or far field electromagnetic effect. Crosstalk is a common case of radiated coupling, defined as interference to a signal path from other localised signal paths, which can be caused by an electric field coupling by a mutual capacitance or a magnetic field coupling by a mutual inductance.

In many cases coupling is unwanted and termed *Electromagnetic Interference* (EMI). Standards bodies, such as the International Standards Organisation, the International Electrotechnical Commission and the European Union, have issued standards citing the maximum permitted emissions from a device and the minimum level of interference under which a device must still correctly operate (Carr 2001, p. 293).

The interest for a security-minded attacker is that these standards assume interference is unintentional and not malicious. The attacker can exceed the permitted level of interference in an attempt to cause malfunction of the security device. Moreover, he can craft the interference to be optimal to attack his chosen device. The device cannot afford to be too sensitive to such injection, otherwise it will be susceptible to EMI and unlikely to operate correctly in a non-hostile noisy environment. Military documents such as NSTISSAM TEMPEST/2-95 (National Security Telecommunications and Information Systems Security 1995) and AFSSM-7011 (United States Air Force 1998) stipulate protections against HIJACK and NONSTOP, but much of this work remains classified and there is little evidence of similar precautions being taken in the non-military context.

Of the two coupling modes, a conducted-field attack can be formed by applying an electrical signal to the wiring of a device or its package. A radiated-field attack can be achieved by creating a high-strength varying magnetic or electric field in proximity to the device. I investigated only injection by conduction, while looking at the radiated field, due to the equipment I had available. Burnside, Erdogan, and Arslan (2008), who co-incidentally performed independent similar experiments, considered

the attack with radiated field injection. They named it the *re-emission side-channel*.

5.2 THE CONDUCTED-FIELD RE-EMISSION ATTACK ON LOCHSIDE

5.2.1 *Frequency injection*

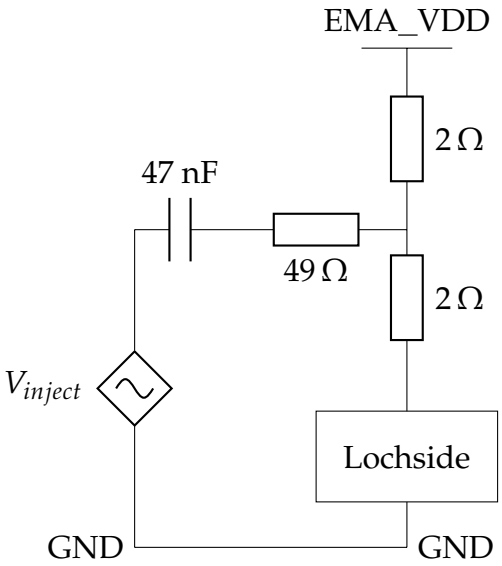
To demonstrate the amplitude modulation effect, I used the ring oscillator on the Lochside test chip. I injected a voltage into the power rails, with the intention of seeing whether an amplitude modulated version could be seen in the electric or magnetic fields. To inject, I inserted resistors into the power rails of the Lochside chip, and coupled in the output from the signal generator (Figure 5.1a on the following page).

The equivalent circuit for AC analysis can be seen in Figure 5.1b on the next page. The power supply is of low impedance to high frequency signals (since it contains a lot of smoothing capacitance), so its impedance $Z_S \approx R_S \approx 0$. R_L, the impedance of the Lochside chip's power rails, is unknown and varying. In this case I chose $R_1 = 49\,\Omega$ and $R_2 = R_3 = 2\,\Omega$. If we assume that $R_L \gg R_2$, most of the AC current flows through R_2 not R_3. If $R_L \ll R_2$, most of the current flows through R_3. The first case is more likely, as there is no low impedance path between the Lochside power rails), and so the voltage across R_2 is approximately:

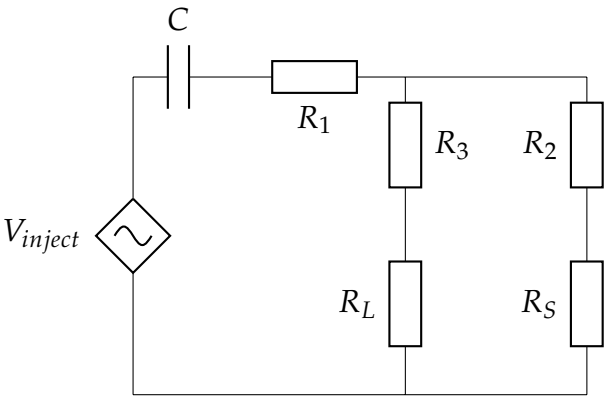
$$V_{R_2} = \frac{R_2}{\frac{1}{j\omega C} + R_1 + R_2} V_{\text{inject}} \tag{5.1}$$

where $j = \sqrt{-1}$ and $\omega = 2\pi \times$ frequency. At $f = 200\,\text{MHz}$, the reactance of C is negligible at $-0.017j$, so we can ignore it. That leaves a simple potential divider: for the values chosen, $V_{R_2} = (2/51)V_{\text{inject}} = 0.039V_{\text{inject}}$.

Thus only a small proportion of the input voltage is delivered across the load; but this method also means the load impedance of the signal generator is approximately matched (at about $51\,\Omega$, depending on the Lochside behaviour).



(a) Injection circuit



(b) Equivalent circuit for AC analysis

Figure 5.1: *Lochside frequency injection*

5.2.2 Amplitude modulation

5.2.2.1 Harmonic distortion and intermodulation

Harmonic distortion may be caused by a non-linearity in the signal path. A non-linear device can be something as simple as a diode junction (intentional or parasitic). When the non-linear device is stimulated by a frequency f_1, it can produce harmonics at frequencies $2f_1, 3f_1, 4f_1, \ldots$.

Intermodulation distortion occurs when signals of two or more different frequencies f_1, f_2, $\ldots f_N$ are presented at the non-linear device. The output contains each of the fundamental frequencies, plus signals at linear combinations of those frequencies.

Considering a two-input system with an input signal:

$$x(t) = A_a \sin(2\pi f_a t + \phi_a) + A_b \sin(2\pi f_b t + \phi_b) \tag{5.2}$$

the result after intermodulation is given by:

$$y(t) = \sum_{m,n \in \mathbb{Z}} A_{m,n} \sin(2\pi(m f_a + n f_b) + \phi_{m,n}) \tag{5.3}$$

where $\mathbb{Z}$ is the set of integers. The coefficients $A_{m,n}$ give the powers of each component, which are derived by Bartlett (1933, 1934). The order of the intermodulation is given by $m + n$. In my experiments I concentrate on second-order intermodulation, in particular at frequencies $f_a + f_b$ and $f_a - f_b$.

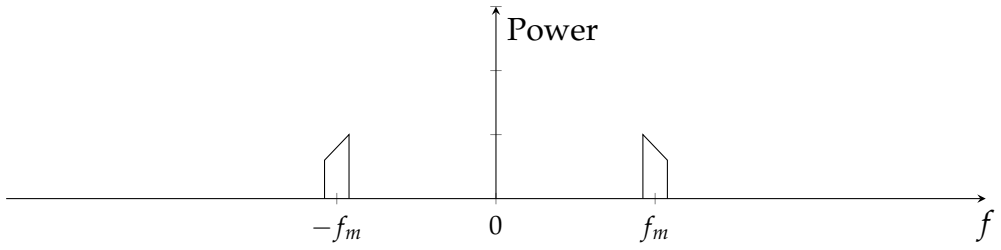
Now consider a system with a time-domain signal $m(t)$ having an operating spectrum of $M(f)$ at natural frequency f. An attacker can inject an additional frequency f_c. Considering second-order intermodulation only and neglecting components apart from $f \pm f_c$, the resulting spectrum is given by:

$$S(f) = \frac{A_c}{2}[\delta(f - f_c) + M(f - f_c) + \delta(f + f_c) + M(f + f_c)] \tag{5.4}$$

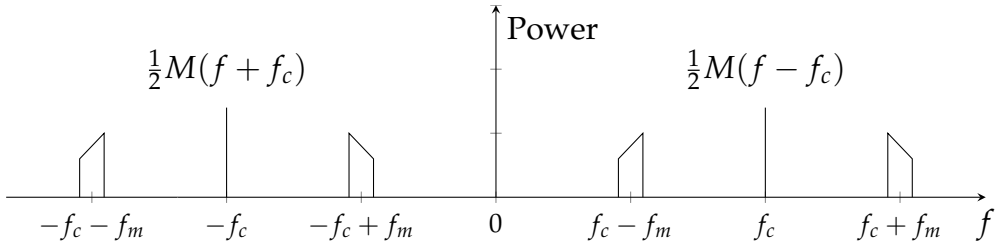
where $\delta(f)$ is the Dirac delta function (Couch 2000, p. 238). This subset of the intermodulation products is equivalent to $m(t)$ being amplitude modulated by injected frequency f_c.

As shown in Figure 5.2 on the following page, two sidelobes are created at $f_c \pm f_m$. Unless it is deliberately filtered out (as in many communication systems), the carrier f_c will also be present.

Thus the signal f_m may be shifted to a higher or lower frequency. This changes its propagation behaviour, and it may enable the attacker to:



(a) Baseband signal $M(f)$



(b) $M(f)$ modulated by carrier of frequency f_c

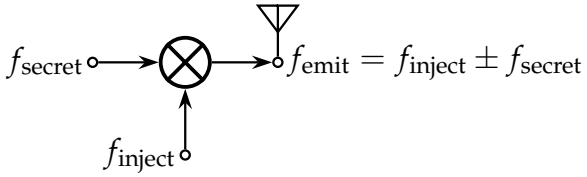
Figure 5.2: *Frequency spectrum of a signal with an arbitrary spectrum centred on frequency f_m, after intermodulation on carrier of frequency f_c*

- receive from a longer distance;
- use a smaller, more convenient or more covert antenna;
- tune the injected frequency to match their receiver;
- increase the signal power; and
- defeat shielding intended for baseband or lower frequency signals.

One of the difficulties in carrying out EM analysis is the need to build a broadband receiver, with an antenna that exhibits a flat response over a wide range of frequencies. In this way we can convert the problem to a narrowband receiver at a higher frequency, and it may be suitable to use a commercial narrowband receiver. Tuning the injected frequency may enable the listener to select different parts of the emitted spectrum for inspection, taking into account any frequency nulls that may exist.

In effect, we create a superheterodyne system, with the secure circuit as the mixer (Figure 5.3 on the next page). The injection frequency can be adjusted to tune the output harmonic(s) to suit the Intermediate Frequency stages.

Frequency injection and emission:



Intermediate frequency receiver:

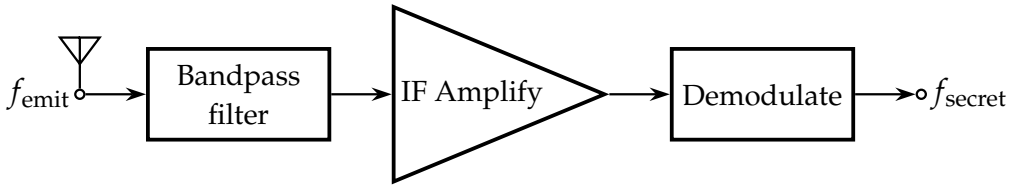


Figure 5.3: *Superheterodyne receiver formed by device injection*

Tuning can move frequencies of interest to those supported by commercial receivers such as those for AM, FM, analogue UHF TV, Digital Audio Broadcast, amateur radio, mobile phones or satellite receivers. The receiver frontend is of more interest than the digital decoding steps aimed at a particular standard (such as DAB or GSM), so the signal can be tapped off after downconversion or the DSP could be reprogrammed to perform the attack.

5.2.3 Measurements

A modulating sinusoid was injected into the 1.8 V core supply for the Lochside chip, using the coupler described in Section 5.2.1 on page 143. The ring oscillators were set running, and the chip lid electric field sensor was used to measure the field under various injection frequencies, with the oscillators either driving a pin or driving no external output.

5.2.3.1 Output pad on

I measured the spectrum with the long ring oscillator not running and no injected frequencies and injected a 200 MHz signal into the core power with the oscillator pad in operation (Figure 5.4 on page 149). 40 mV was applied across the Lochside chip. Subtracting the first spectrum

from the second, I saw a clear 200 MHz carrier harmonic and sidelobes at 200±6.2 MHz. Therefore I was modulating up the (approximately) 6.2 MHz ring oscillator frequency.

5.2.3.2 *Output pad off*

I took similar measurements, this time with the output pad turned off, to test whether, by this technique, I could see a signal passing through only 2597 transistors with no pad activity.

I first tried using the long ring oscillator, as above. No obvious cross-modulated signals could be seen in the E-field.

Then I tried running the short ring oscillator with its output turned off, and injecting a harmonic into the core V_{DD}. The short ring oscillator's output f_c was identified at 1121 MHz by E-field measurements with the pad on and off. Then a 25 MHz signal f_m at 82 mV was injected into the V_{DD} rail. Figure 5.5 on page 150 depicts the effect of the high frequency ring oscillator being amplitude modulated by the lower injected frequency. The $f_c + f_m = 1121 + 25$MHz and $f_c - f_m = (1121 - 25)$ MHz sidebands from amplitude modulation can be clearly seen.

This particular example, of shifting around a 1 GHz signal by 25 MHz, is not terribly useful but it does demonstrate the principle. The problems with exploring it further were that I could not alter the internal oscillator frequency (apart from with the supply voltage, which would have disturbed the measurement). I also had not tested my antennas at such high frequencies (the available signal generator was limited to 240 MHz) so I had nothing to benchmark the results with except the oscillator under test. This meant there were too many unknown variables for such a test to be usefully expanded, but it demonstrated it was possible.

5.3 FREQUENCY MODULATION

Signals can be frequency modulated in a number of ways. The simplest is a coupling between the supply voltage and some timing-related element. This might be the free-running frequency of a ring oscillator, or the computation time of an asynchronous circuit.

To demonstrate this, I looked at the response of the Distributed Clock Generator (DCG) on the Lochside chip to an injected signal on the power supply. The DCG's frequency is controlled by an analogue control voltage provided on an external pin. Figure 4.35 on page 123 gives a graph of frequency against input voltage, where the useful operating region is about 0.2 to 0.9 V.

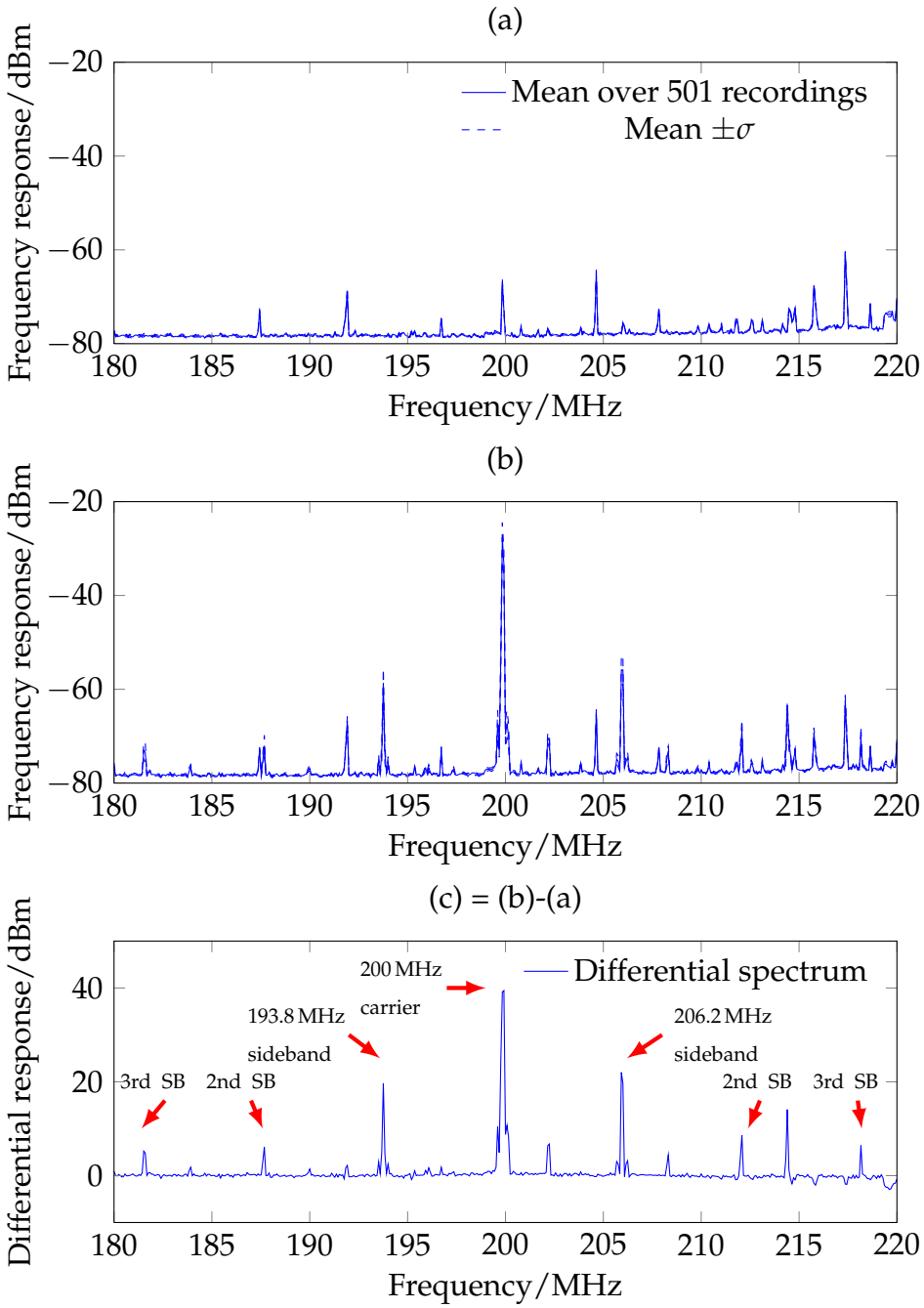


Figure 5.4: *Lochside E-field with clock output pad turned on. (a) Long ring oscillator off, no frequency injection into power supply. This shows the background radiation received. The frequency generator was connected with its output off, but a weak 200 MHz signal does leak out; (b) Long ring oscillator on, inject 200 MHz at 40 mV. As well as the main carrier peak, new peaks at 193.8 MHz and 206.2 MHz appear; (c) Differential spectrum shows the harmonics more clearly.*

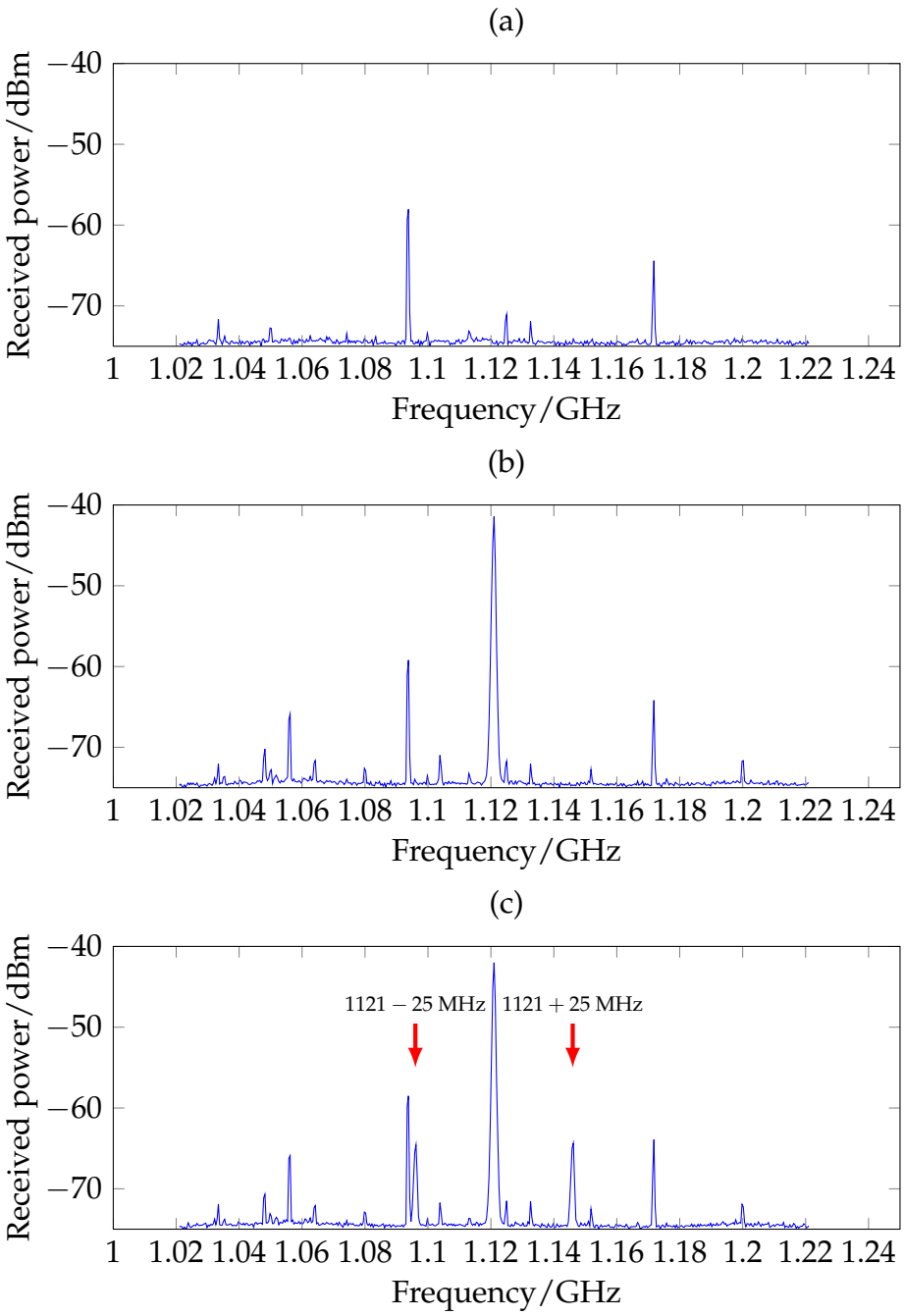


Figure 5.5: *Lochside E-field with output pad turned off: (a) Ring oscillators not running, no injected frequency; (b) short ring oscillator running, no injected frequency; (c) short ring oscillator ring, inject 25 MHz at 82 mV into power supply. Additional harmonics at 1096 and 1146*MHz *show amplitude modulation of the low frequency power spectrum onto the higher frequency of the oscillator.*

Thus it is possible to turn the DCG into a frequency modulator very simply, by applying a DC level plus modulating frequency to the control pin. For this test, however, I held the control pin at a fixed voltage with a low impedance bench power supply; a short, thick wire; and a 100 nF capacitor to ground on the board. Any modulating influence will thus have to be via the on-chip power supply rather than AC signals induced onto the control voltage.

In this way the existing high-frequency signal on the board can be frequency modulated by a power supply signal. This is effectively modulating the power trace (as used in DPA) onto a high frequency carrier.

The Lochside control voltage was set to 711 mV and the DCG output was turned on, which produced a clear 181 MHz harmonic, detectable in the E-field. Through the 50:1 coupler a 2 MHz signal was injected into the power supply, at strengths varying from 2 to 83 mV (from 50 mV to 2 V peak-peak before the coupler, in 100 mV intervals).

The spectrum of a frequency modulated signal is not easy to determine analytically, but can be derived if the applied frequency is a sinusoid. According to Couch (2000, p. 324), the frequency spectrum for $g(t) = A_c \cos \omega_c t$ is given by:

$$G(f) = A_c \sum_{n=-\infty}^{n=\infty} J_n(\beta)\delta(f - nf_c) \tag{5.5}$$

where $J_b(\beta)$ is the Bessel Function of the First Kind of the n-th order (to be found from tabulated values). Modulation index β for FM is defined by $D_f A_c/\omega_c$, where constant D_f is the furthest distance the output frequency may deviate from carrier f_c, a function of the gain of the VCO, with units of radians/volt-second. For example, in FM radio broadcast at 88 to 108 MHz, the peak frequency deviation D_f is 75 kHz. Thus the n-th harmonic of f_c can expect a magnitude $A_c J_n(\beta)$.

With larger values of β we see weaker but broader spectra (Figure 5.6 on the following page).

A sample of the spectra measured from the Lochside E-field can be seen in Figure 5.7 on page 153. The distribution of the spectral peaks with increasing injected power are also displayed in an intensity map in Figure 5.8 on page 154. The same plotting style is used to display the Bessel function in Figure 5.9 on page 155. From this it is easy to recognise that FM is taking place, with $\beta \approx 0$ to 1.5.

Given I only had access to a scalar spectrum analyser, the phase is missing from these recordings so it is not possible to verify them by direct demodulation. This could have been achieved by taking a set of time domain

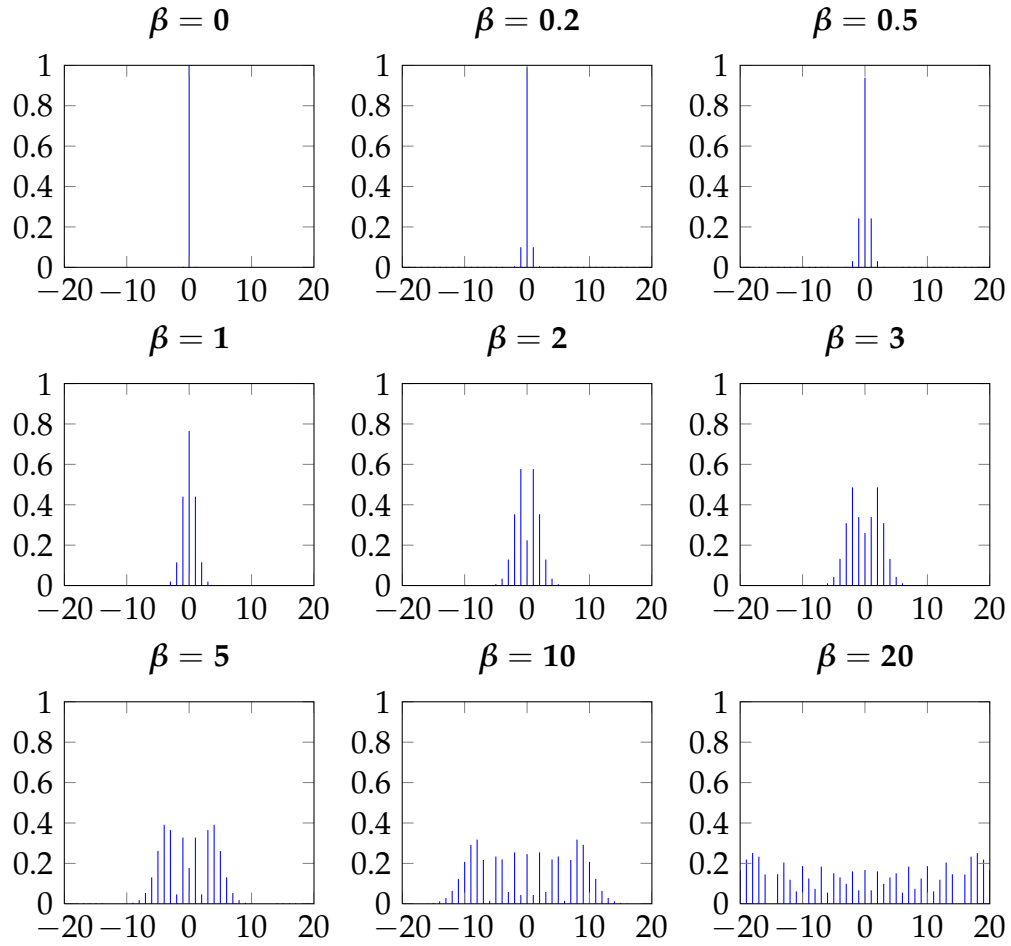


Figure 5.6: *Bessel Functions of the First Kind, the order b on the horizontal axis. In FM, spectra are centred at f_c (order 0) and each harmonic $f_c + bf_m$ is of order b. All Bessel peaks would be scaled by A_c.*

measurements which I did not have time to do.

5.4 SPRINGBANK RE-EMISSION ATTACKS

5.4.1 *Springbank frequency-domain power analysis*

The simplest way to demonstrate re-emission is to look at data-dependent frequency components. I have described in Chapter 4 the data-dependent timing effect of the Secure XAP on Springbank; now I aimed to detect this in the frequency domain.

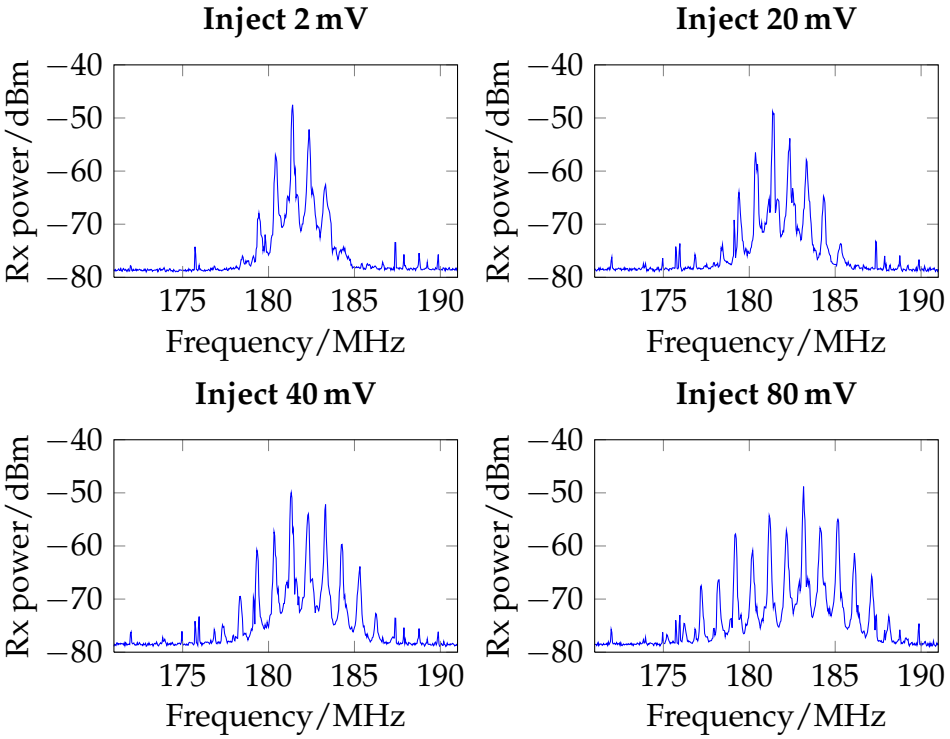


Figure 5.7: *E-field spectra from Lochside: DCG running with fixed control voltage, injecting 2 MHz into power supply. The effect correlating to FM can be clearly seen (compare with Figure 5.6 on the preceding page).*

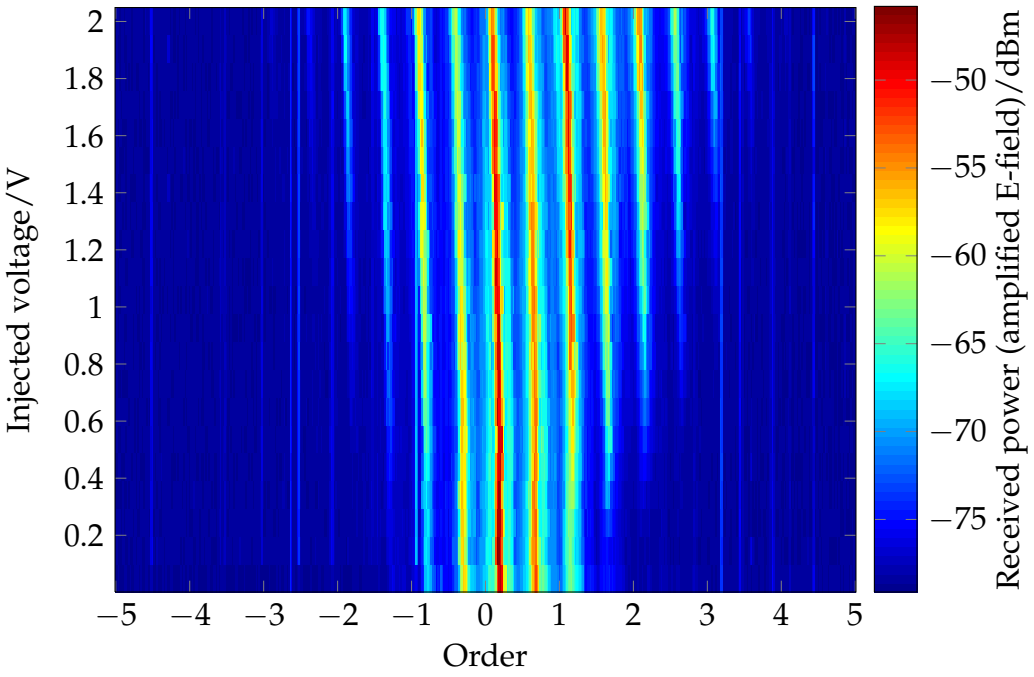


Figure 5.8: *Spectral peaks of Lochside FM, from same configuration as Figure 5.7 on the previous page. Increasing injected voltage is displayed along the vertical axis, with the frequency spectrum at that voltage showing as a stripe in the horizontal direction. The carrier is centred at order 0 (181 MHz), and the harmonics (one order point=2 MHz) are visible. The spectra become more spread with higher injected powers. Note this figure is interpolated from 21 spectral stripes.*

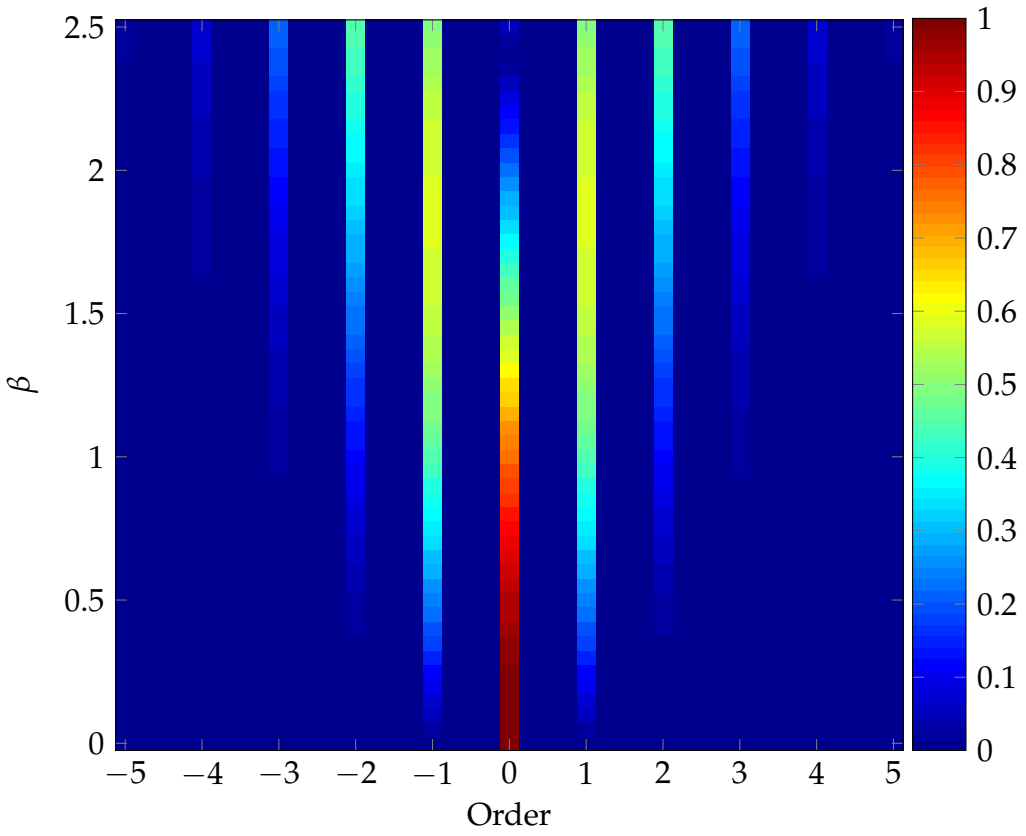


Figure 5.9: *Bessel Functions of the First Kind displayed in the same plotting style as Figure 5.8 on the preceding page. For clarity each vertical stripe (being a single infinitely sharp frequency) is given an arbitrary finite width.*

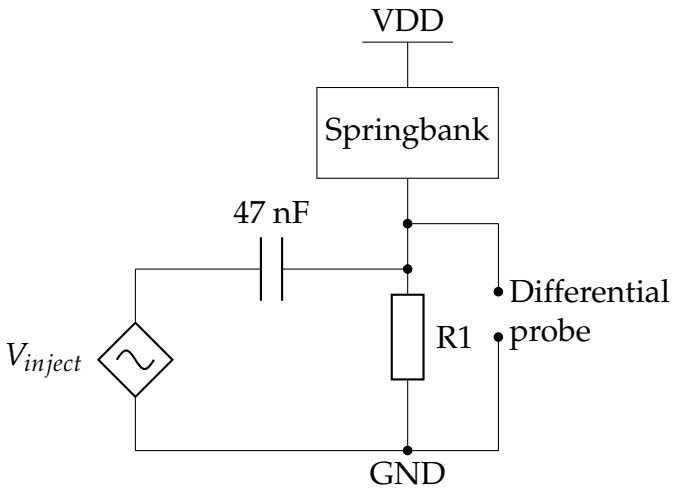


Figure 5.10: *Springbank power analysis and frequency injection circuit. For frequency injection, R1 is 10 Ω and the differential probe is omitted.*

I used the test program `XAP_XOR00to`*XX*`NoOutput` where *XX* = 00 or FF; it has a very small inner loop:

```
itloop_dotest:  ld      ah,#H'00
                xor     ah,#H'FF
                bra     itloop_dotest
```

or

```
itloop_dotest:  ld      ah,#H'00
                xor     ah,#H'00
                bra     itloop_dotest
```

In a system with data-dependent timing, the natural frequencies of the two loops may be slightly different and this should be visible in the spectrum. As with all XAP code, while there are four registers, it is impossible to do a logical operation between two registers as one operand must always come from memory. Thus we are performing an attack on a combination of ALU and on-chip memory. This is typical of microcontroller architectures.

A 10 Ω resistor was inserted in the core VDD line of the Springbank chip, and the oscilloscope differential probe inserted as seen in Figure 5.10.

Taking an overall spectrum 0 to 100 MHz, we spot small differences in peak frequencies around 37 MHz and 62 /68 MHz (Figure 5.11 on the next page). Taking a much more magnified spectrum of the 37 MHz peak, we find a detectable difference in frequencies (Figure 5.12 on page 158).

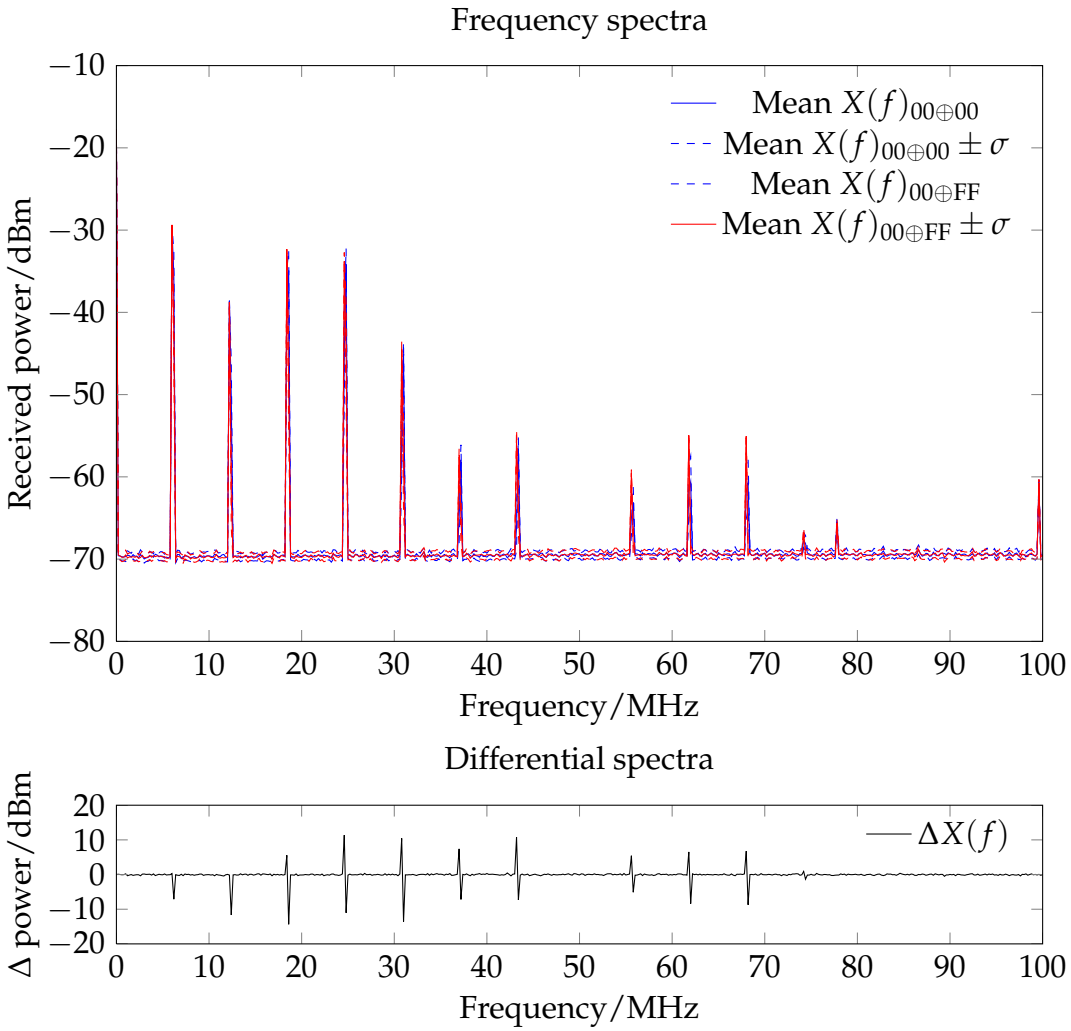


Figure 5.11: *Power consumption spectrum of Springback running* `XOR00to-XXNoOutput`. *Small differences can be seen at 37, 62 and 68 MHz, and at other points*

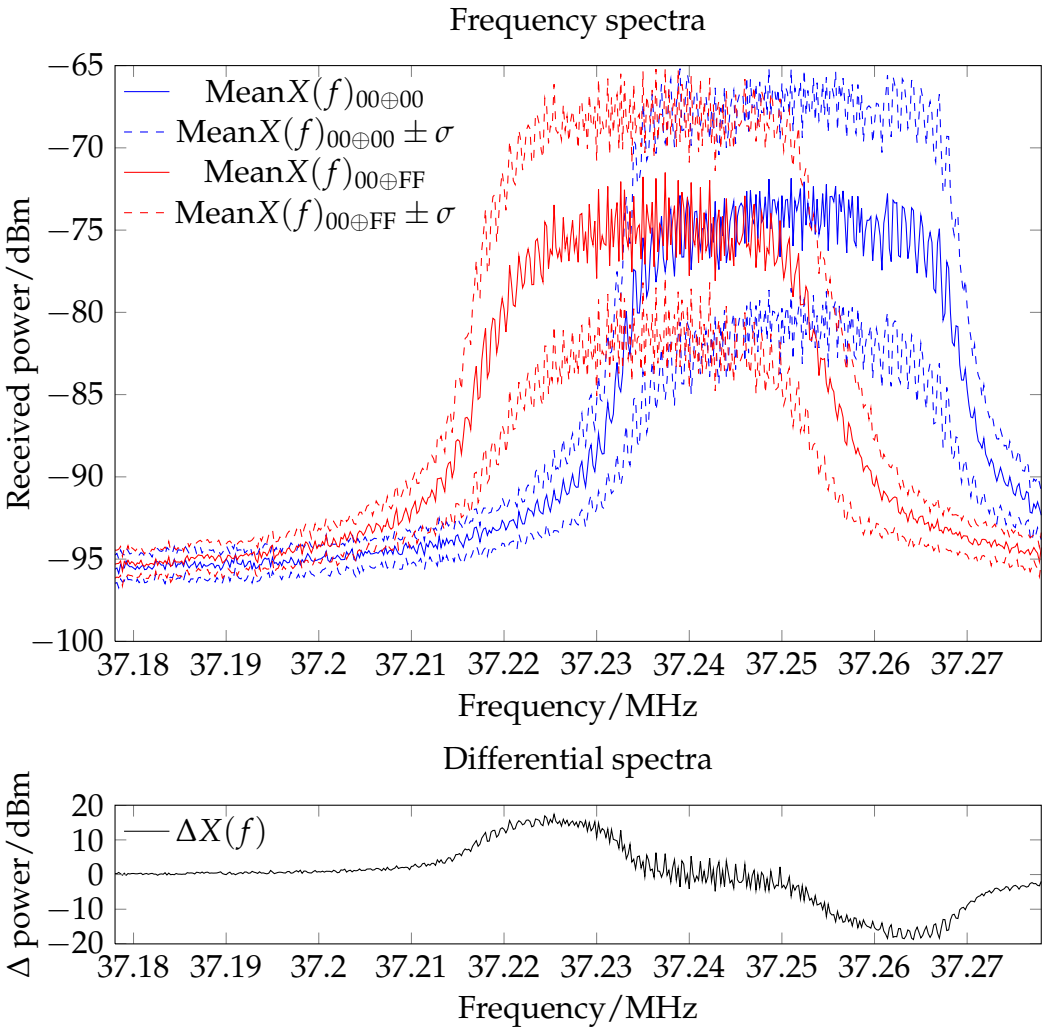


Figure 5.12: *Springbank data-dependent baseband spectral shift in power consumption: a higher-resolution recording of part of Figure 5.11 on the preceding page. A 20 kHz shift was noted at 37.25 MHz.*

5.4.2 Re-emission analysis

The electric field emissions of the Springbank chip were measured with a chip lid E-field sensor, similar to that used on the Lochside chip (as seen in Figure 4.5 on page 83). The injection circuit of Figure 5.10 on page 156 was used to apply a varying ground rail of the Springbank chip. The input was not correctly terminated to avoid severe attenuation or disruption of the Springbank DC power supply arrangement, which gives some variability in injected powers with frequency.

The same spectrum as Figure 5.12 on the preceding page, measured in the E-field, may be found in Figure 5.13 on the following page. The E-field peaks are sharper, but the same relationship can be seen.

When a 200 MHz carrier at 280 mVpk−pk is injected into the power supply, there are plenty of frequencies visible around the carrier, as may be seen in the 0 to 450 MHz plot Figure 5.14 on page 161.

There appear to be some data-dependencies, though some are in fact discrepancies due to the limited horizontal sampling resolution of the spectrum analyser.

To test re-emission, a carrier frequency of 200 MHz was injected into the power supply. Figure 5.11 on page 157 indicates that, in baseband mode, there is a data leakage in the 25 MHz region. This region was examined, both in baseband (25 MHz) and the upper sideband at $200 + 25$ MHz.

Figure 5.15 on page 162 shows the 24.7 MHz peak in more detail: a shift in frequency of about 10 kHz is visible from the data-dependency.

Figure 5.16 on page 163 shows the upper sideband at 224.7 MHz; whilst weaker and more noisy, the same relation can be heard. Thus the harmonic which leaks information has been up-converted.

This frequency lies within the UK 217.5 to 230 MHz Digital Audio Broadcast (DAB) range (Emery undated, National Frequency Planning Group 2008), and so a conventional DAB antenna can receive it. Either a DAB receiver may be modified to tap off the tuned signal before analogue to digital conversion, or a more general amateur radio receiver (a 'scanner') may be used.

5.5 Conclusion

In straightforward logic circuits, I have demonstrated that it is possible to up-convert internal operation signals into others which are detectable in the electric field of the chip.

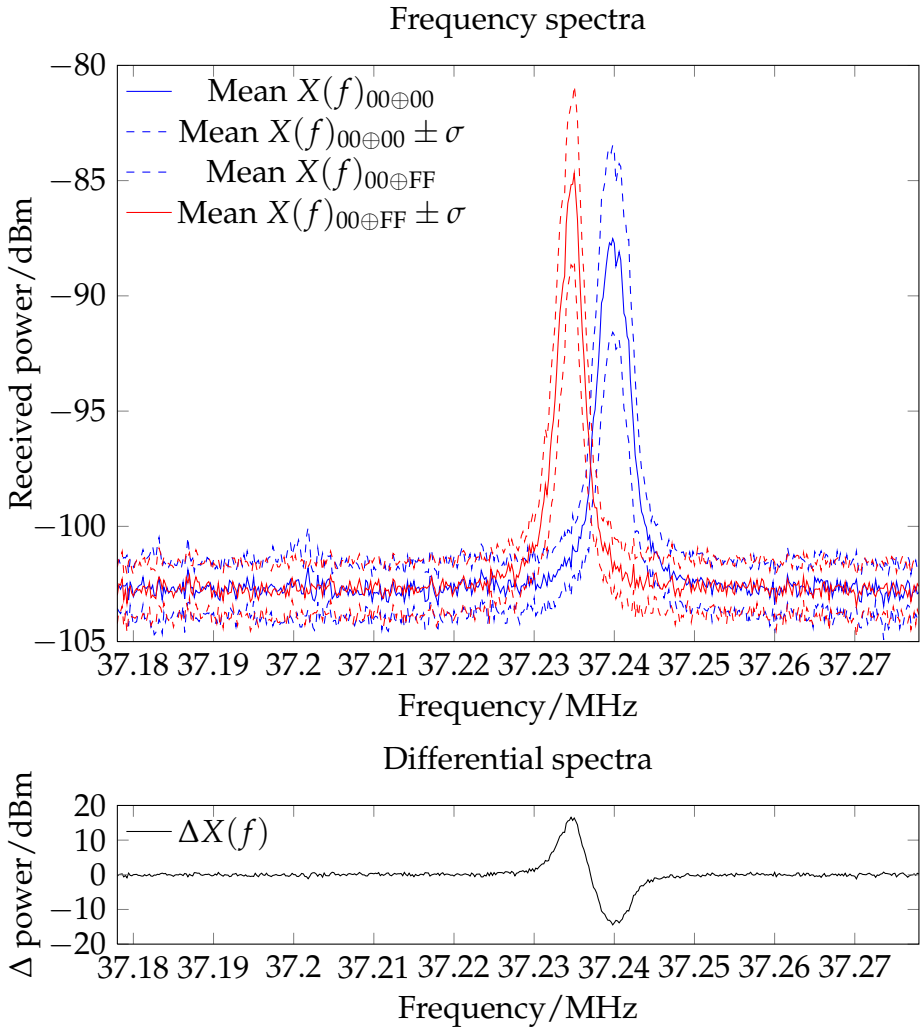


Figure 5.13: *Same recording conditions as Figure 5.12 on page 158, but measuring electric field. This trace shows a similar baseband data-dependent frequency shift, but slightly smaller.*

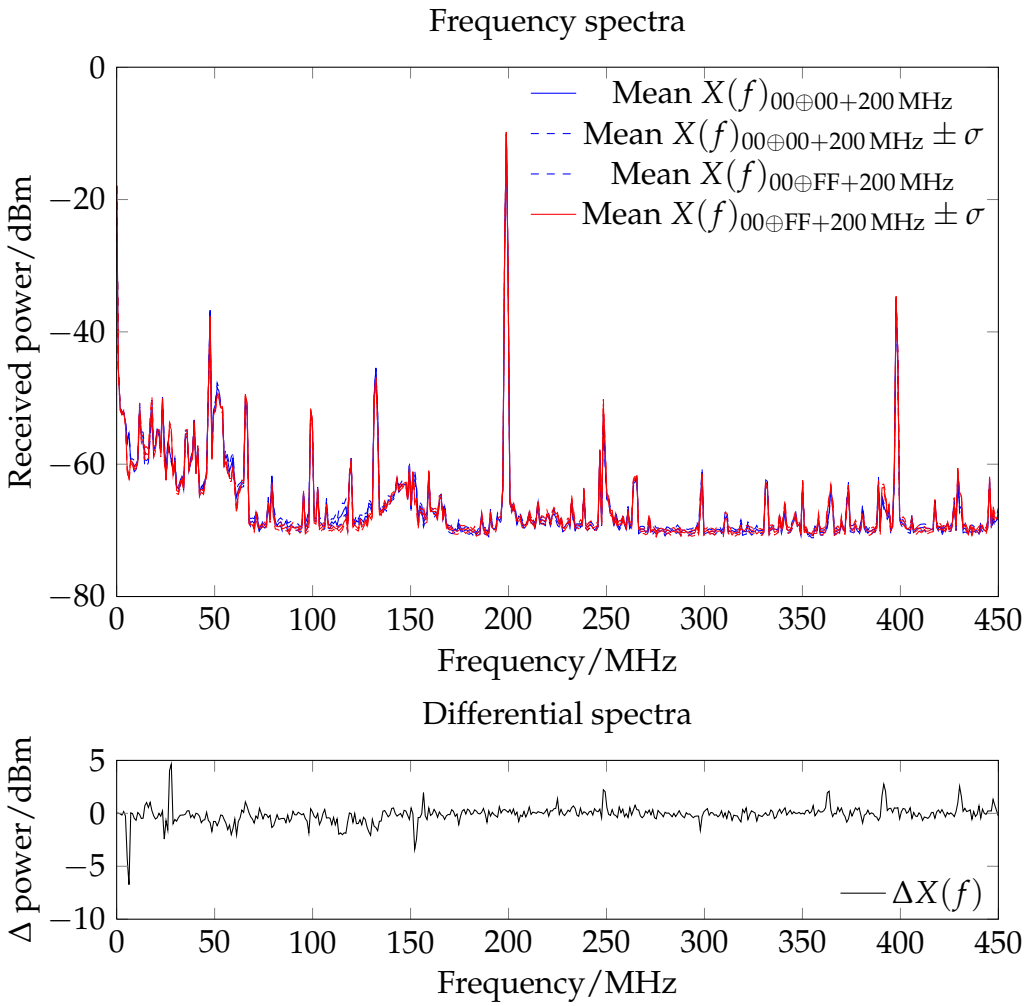


Figure 5.14: *Spectrum from E-field sensor on Springbank running* `XAP_XOR-00toXXNoOutput` *with 200 MHz carrier at 280 mVpk–pk injected into the power supply. Suspected data-dependencies detected at various points in the spectrum, notably in the 25 MHz area. The peaks at 25 MHz and 225 MHz are reproduced in the following figures.*

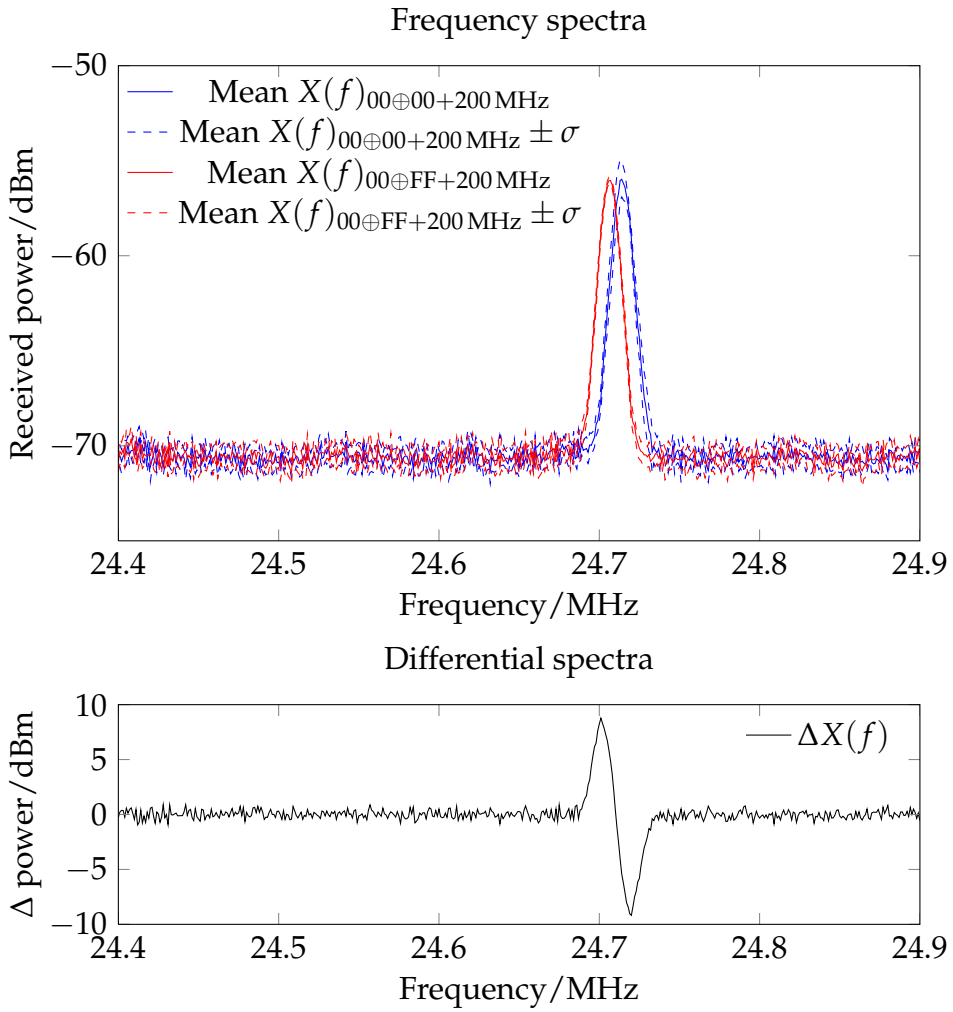


Figure 5.15: *Data-dependent harmonic at 24.7 MHz, from the same conditions as Figure 5.14 on the preceding page*

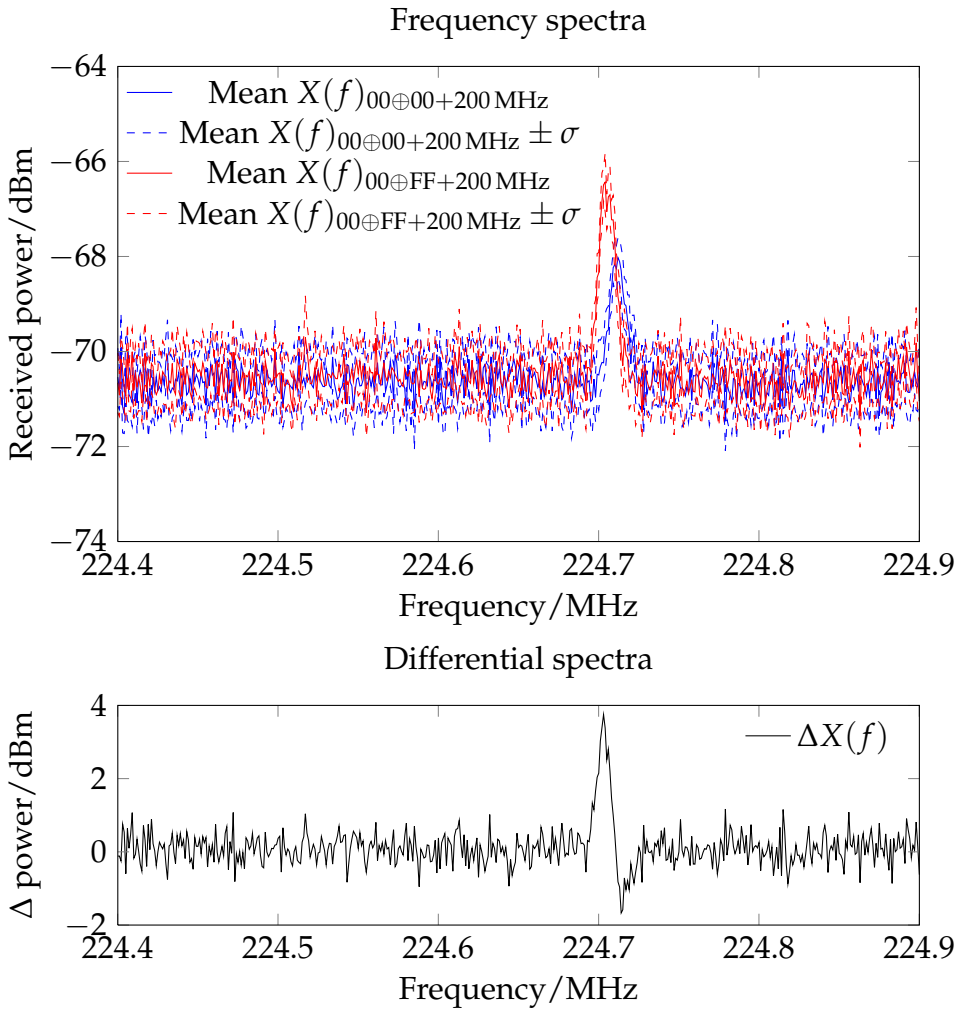


Figure 5.16: *Up-converted data-dependent harmonic at 224.7 MHz due to 200 MHz frequency injection, from the same conditions as Figure 5.14 on page 161*

oscillators.

In the frequency modulation case, I used an asynchronous ring, and I was able to affect its frequency of operation by injecting into the power supply. A very clear demonstration of frequency modulation was visible for modulation indicies $\beta \approx 0$ to 1.5.

Having tested the principle in the limited field of oscillators, I used the re-emission attack on a secure microcontroller to broadcast data-dependent timing in both power and electric field emissions.

These showed the promise of the technique in ad hoc experiments. I was intending to perform more detailed and rigourous experiments when I learnt of the results of Burnside, Erdogan, and Arslan (2008). They used both conducted and radiated injection rather than the conducted injection that I used. Radiated injection is a more realistic attack scenario but less repeatable in the laboratory (plus they admit that a major source of EM coupling its through wiring). They also had more advanced results, by attacking DES on a commercial smartcard, and by considering an active null re-emission technique to prevent large injection frequencies swamping small data signals. Both of these I had intended to cover, but felt that I would not be adding significant new material to Burnside's work.

Instead I focused on cryptographic random number generators. As we shall see in the next chapter, the vulnerability is different but similar techniques apply.

CHAPTER 6
THE FREQUENCY INJECTION ATTACK ON RANDOM NUMBER GENERATORS

6.1 INTRODUCTION

Random numbers are a vital part of many cryptographic protocols. Without randomness, transactions are deterministic and may be cloned or modified. Some devices generate their random numbers by measuring jitter in ring oscillators. What if it were possible to use my frequency injection techniques to modify the jitter being used to collect randomness? This would present a new non-invasive attack.

Consider an example in the EMV banking protocol. For cash withdrawal an Automatic Telling Machine (ATM) picks an unpredictable number from four billion possibilities. Imagine if an insider can make a small covert modification to an ATM to reduce this to a small number, R. This might not need insider access. He could then install a modified EMV terminal in a crooked shop. A customer enters and pays for goods on their card. While the modified terminal is doing the customer's EMV transaction with their secret PIN, it simulates an ATM by performing and recording $\sqrt{R}$ ATM transactions. The customer leaves, unaware that extra transactions have been recorded.

The crooked merchant then takes a fake card to the modified ATM. The ATM will challenge the card with one of R random numbers. If the shop recorded a transaction with that number, he can withdraw cash. If not, the fake card terminates the transaction (as might happen with dirty card contacts) and starts again. By the Birthday Paradox I only need roughly $\sqrt{R}$ attempts at the ATM before, at least, a 50% chance of success. The customer has no defence: both their card and PIN were used in the transaction, just not at the time they expected.

In this attack I have reduced the ability of a microcontroller known to be used in ATMs to produce 4 billion (2^{32}) random numbers, to just 225 ($< 2^8$). For more than a 50% chance of a successful attack, I only need to record 13 transactions in the shop and try 13 transactions at the ATM. My attack is based on injecting signals into the power supply of a smartcard

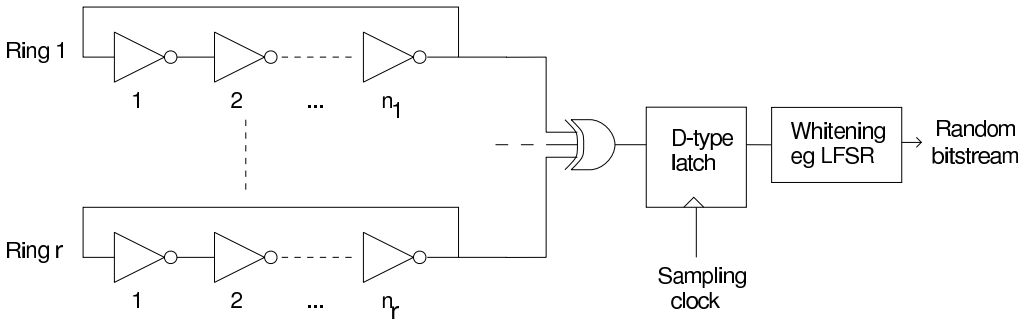


Figure 6.1: *Outline of the basic ring oscillator TRNG.*

or secure microcontroller. It is made possible by adding a small number of extra components costing tens of dollars.

6.2 RANDOM NUMBER GENERATION

A *true random number generator* (TRNG) must satisfy two properties: possessing *uniform statistics* and containing a source of *entropy*. Non-uniform statistics might enable the attacker to guess common values or sequences. Entropy comprises a source of uncertainty in a normally predictable digital system. Failure of these properties in even subtle ways leads to weaknesses in cryptographic systems (Bellare, Goldwasser, and Micciancio 1997, Bello 2008).

A common implementation of a TRNG is provided by comparing free-running oscillators. These are designed to be sensitive to thermal, shot or other types of random noise, and present it as timing variations. Such timing variations can be measured by a digital system and the entropy collected. An oscillator that is easy to fabricate on a CMOS digital integrated circuit is the ring oscillator (see Figure 6.1) which is used in many TRNG designs.

Practical TRNG sources are typically *whitened* by post-processing before cryptographic use to ensure uniform statistics. Typically whitening functions include calculating the remainder of a polynomial division using a linear-feedback shift register (LFSR) or hash functions. If the entropy source is removed, TRNG outputs revert to a repeating sequence from the whitening function.

In this chapter I examine the operation of the ring oscillator and explain how the principle of *injection locking* may be used by an attacker to take control of this entropy source.

6.3 THEORY

6.3.1 Ring oscillator TRNG operation

Hajimiri, Limotyrakis, and Lee (1999) give the frequency of a single-ended[1] CMOS ring oscillator formed from N inverters with equal-length NMOS and PMOS transistors to be:

$$f_0 \equiv \frac{\omega_0}{2\pi} \approx \frac{\mu_{\text{eff}} W_{\text{eff}} C_{\text{ox}} (\frac{V_{\text{DD}}}{2} - V_{\text{T}})}{8\eta N L q_{\max}} \tag{6.1}$$

This relates the fundamental frequency f_0 to the gate-oxide capacitance per unit area C_{ox}, the transistor length L, the power supply voltage V_{DD}, the gate threshold voltage V_T and the proportionality constant $\eta \approx 1$. $q_{\max}$ is the amount of charge a node receives during one switching period. I consider both NMOS and PMOS transistors together, giving effective permeability μ and transistor width W:

$$W_{\text{eff}} = W_{\text{n}} + W_{\text{p}} \tag{6.2}$$

$$\mu_{\text{eff}} = \frac{\mu_{\text{n}} W_{\text{n}} + \mu_{\text{p}} W_{\text{p}}}{W_{\text{n}} + W_{\text{p}}} \tag{6.3}$$

These are all physical constants determined in the construction of the oscillator. A ring oscillator with no other effects would be completely predictable.

Oscillators do not have a perfectly stable output. In the time domain, random noise means they sometimes transition before or after the expected switching time (Figure 6.2 on the following page). In the frequency domain, this implies small random fluctuations in the phase of the wave, slightly spreading its spectrum and adding noise to the phase of the signal (Figure 6.3 on the next page). This same effect is referred to as *jitter* in the time domain and as *phase noise* in the frequency domain. They are both cumulative over time.

Phase noise can be represented by adding a random noise component $n(t)$ to the phase of the signal, i.e. for a sine wave generator:

$$v(t) = A\cos(\omega t + n(t)) \tag{6.4}$$

[1] In a *single-ended* ring the connection between each node is a single unbalanced signal, as opposed to a *differential* ring, in which each connection is a balanced pair.

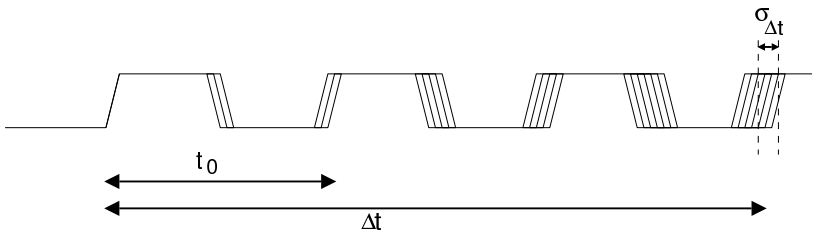


Figure 6.2: *Jitter in the time domain causes increasing uncertainty in the timing of transitions.*

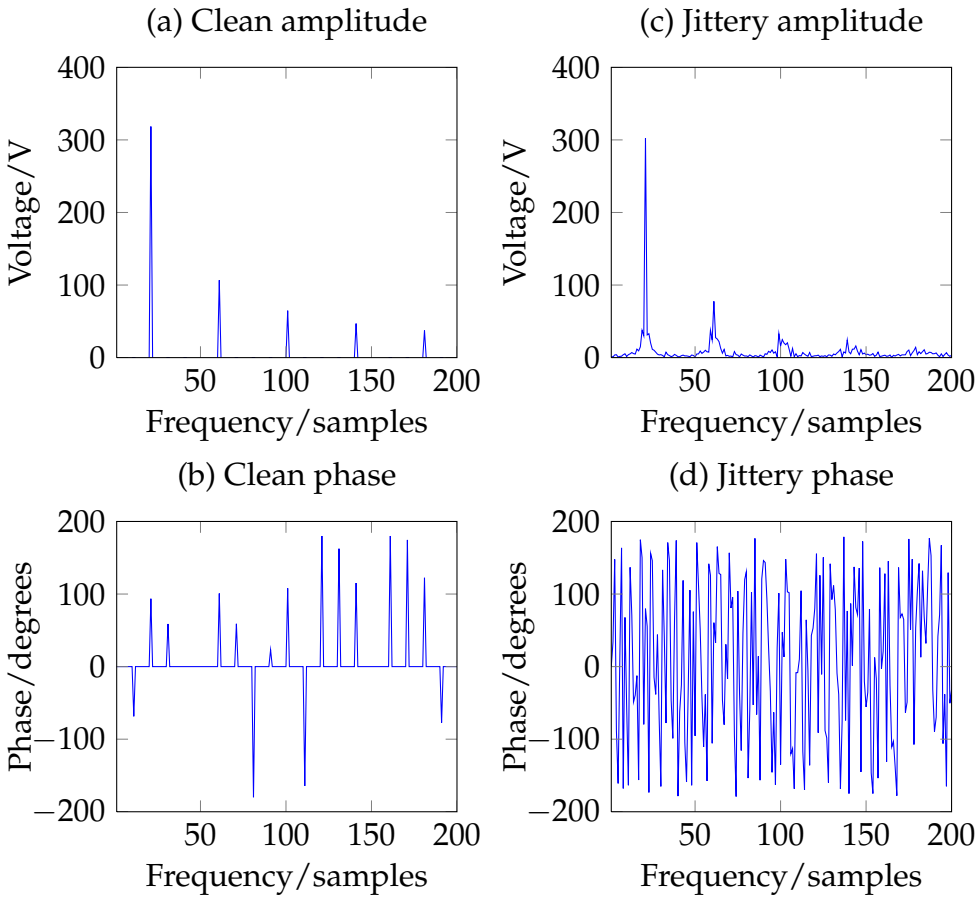


Figure 6.3: *Frequency spectrum of jitter of 20 cycles of a simulated ±0.5 V square wave with period 50 samples, both with and without a normally distributed jitter component of ±1 samples ($\sigma \approx 1.04$) per transition. The frequency of harmonics is spread a little by the jitter, while noise appears on the phase.*

In a single-ended ring oscillator, at a time Δt after the starting, Hajimiri derives that the jitter due to thermal noise will have a standard deviation:

$$\sigma_{\Delta t} \approx \sqrt{\frac{8}{3\eta}} \sqrt{\frac{kT}{P} \frac{V_{\mathrm{DD}}}{V_{\mathrm{char}}}} \sqrt{\Delta t} \tag{6.5}$$

where P is the power consumption and kT the Boltzmann constant multiplied by temperature. V_{char} is the characteristic voltage across a MOSFET; in the long-channel mode it is $\frac{2}{3}((V_{\mathrm{DD}}/2) - V_{\mathrm{T}})$.

This is equivalently written as a phase noise spectrum:

$$L\{\omega\} \approx \frac{8}{3\eta} \frac{kT}{P} \frac{V_{\mathrm{DD}}}{V_{\mathrm{char}}} \frac{\omega_0^2}{\omega^2} \tag{6.6}$$

where ω_0 is the natural angular frequency of the oscillator and variable ω is some deviation from it (i.e. $\omega = 0$ at ω_0).

In a TRNG based on ring oscillators, jitter is converted into entropy by measuring the timing of transitions: jitter causes the exact timing to be unpredictable. There are two main ways to construct such a TRNG: with relatively prime ring lengths (Eastlake, Schiller, and Crocker (2005) and several patents) and with identical ring lengths (Sunar, Martin, and Stinson 2007). Both employ a topology based on that of Figure 6.1 on page 166. The combined signals from the rings are sampled at some uncorrelated frequency, producing a stream of bits, which is then whitened before cryptographic use.

In the identical rings context, I have two or more rings running at the same frequency. Entropy is wasted when jitter from one transition overlaps jitter from another, since only one transition is measured. Sunar, Martin, and Stinson (2007) extends this to tens or hundreds of rings to increase the probability that at time t there will be a ring R that does not transition. Cumulative jitter is measured as the phase drift between each ring.

With relatively prime rings, the outputs slide past each other, minimising the likelihood of two rings transitioning together. Transition timing is based on a prime factor and the integral of past jitter. Sunar points out that fabrication of relatively prime rings to produce more concentrated entropy is expensive. In my experimental work I concentrated on relatively prime rings, since, I suggest, these are more difficult to lock to an input frequency (or frequencies). For identical rings it should be much simpler.

6.3.2 Frequency injection attacks

Bak (1986) describes how a dynamical system will, at certain frequencies, resonate and, at others, be chaotic. A resonator, such as a pendulum, with resonant frequency m, is capable of locking onto any driving frequency n forming a rational m/n. Given an infinity of such locking frequencies, as n is slowly changed the pendulum will jump from one rational resonant frequency to another. A plot of n against m will show an infinity of fractal steps as the resonator hops from one resonant state to the next – the 'Devil's Staircase'. Similar plots can be seen in many widely different physical, chemical, biological and other systems.

The effect is known as *injection locking*. It was first discovered by Christiaan Huyghens in 1665. He describes (Huyghens 1673, p. 19) how two pendulum clocks on a wall can be heard to swing together and, if disturbed, tend to fall into synchrony. Adler (1946) describes the conditions for lock as applied to a vacuum tube LC electronic oscillator.

My attack constitutes injecting a signal of frequency $f_i := \omega_i/2\pi$ and magnitude V_i into the ring oscillators, causing them to lock to the injected frequency. Locking is a steady state: at lock the relative phase ϕ between the two oscillators is constant, so $d\phi/dt = 0$. Once lock has been achieved, the ring's natural frequency is irrelevant; jitter in the injecting signal will be received equally by all the rings, impairing any TRNG that compares jitter between oscillators.

Mesgarzadeh and Alvandpour (2005) analyse this for a three-ring C-MOS oscillator deliberately made asymmetric by the forcing input being an additional gate overdriving one signal. They prove Adler's work also applies to their ring oscillator. Rearranging their condition for lock in Adler's form, I have:

$$2Q \left| \left(\frac{\omega_i}{\omega_0} - 1 \right) \right| < \frac{V_i}{V_0} \tag{6.7}$$

where V_0 is the amplitude of the oscillator at its natural frequency and Q is its quality factor, a measure of the damping of the oscillator. From my experiments, Figure 6.4 on page 173 shows the difference between rings sliding past each other and in lock.

To achieve injection locking, I must ensure my interference can reach the ring oscillators in a secure circuit. I achieve it by coupling the injection frequency on to the power supply of the device.

The difficulty in proceeding with an analytic solution is determining Q. Adler originally derived the formulation in the context of an LC tank that has a natural sinusoidal operating frequency. Such an oscillator converts energy between two forms, voltage and current in this case. It synchron-

ises to an external signal of suitable frequency to maximize the energy extracted from this driving source. For instance, a pendulum will adjust phase so an external periodic displacement will accelerate its own velocity.

A ring oscillator lacks a clear system-wide change between two alternating states, being just a circle where a rising and a falling edge chase each other, without any natural point defining where a new cycle starts. An idealised 3-element ring consists of three identical inverters, connected via three identical transmission lines. All three inverters and transmission lines oscillate in exactly the same way, but 120° out of phase. A waveform applied via the power supply or an EM field is a *global stimulus* that affects all three inverters equally. It will, therefore, encourage the ring oscillator to synchronise simultaneously with three versions of the stimulus, all 120° apart in phase. Their synchronising effect is thus largely cancelled out.

A global stimulus can only be effective if the three parts are not exactly identical. In a real-world ring oscillator, layout asymmetries, device variations, and loading due to the output tap all break this 120° symmetry, and will allow one of the 120° alternatives to win over the other two. How quickly the ring will lock on to a global stimulus may largely depend on the size of this asymmetry.

Unlike pendula or LC tanks, ring oscillators are also non-linear. In short rings, such as $N = 3$, each gate is in a constant state of transition, so operates linearly, and the output more clearly resembles a sinusoid. But in longer rings, where $N \gg 10$, each gate spends a small fraction of the time in transition, so the ring output is more like a square wave. Adler's model fits this case less well.

6.3.3 *Effect of injection on jitter*

Injection locking can reduce jitter. Consider an oscillator with frequency ω_n being injection-locked from another source. The free-running oscillator will naturally exhibit jitter. If it drifts slightly away from the injected frequency ω_i, it will be slightly disturbed. If the disturbance remains within locking range, injection locking will pull it back to the injection frequency. Indeed, such a technique may be used for clock recovery (Ng, Farjad-Rad, Lee, Dally, Greer, Poulton, Edmondson, Rathi, and Senthinathan 2003).

Considering the output phase, Mesgarzadeh and Alvandpour indicate their injection model can operate on the phase as a first-order low-pass filter with a single pole located at:

$$p = 2\pi\omega_i \ln \frac{1}{1+S} \tag{6.8}$$

where S is the injection ratio V_i/V_0 or, in power terms, $\sqrt{P_i/P_0}$. That is, the rapid variations in phase (phase noise seen in Figure 6.3 on page 168) are filtered out. In the Laplace (or complex frequency) domain such a filter has a transfer function of:

$$H(s) = \frac{1}{1+sp} \tag{6.9}$$

where the Laplace transform operator $s = j\omega$ and $j = \sqrt{-1}$. Its impulse response $h(t)$ is derived by taking the inverse Laplace transform of $H(s)$.

If it were represented as an RC filter on the phase, it would have time constant p. At the cutoff frequency $f_c = 1/2\pi p$, the output magnitude of the phase noise is reduced by 3 dB compared to the input.

There are two effects that may work in concert here. First, the negative feedback effect of injection locking filters out some of the phase noise that is the source of randomness (equation 6.8). In a single oscillator, as in a clock recovery circuit, short term variations of phase are removed.

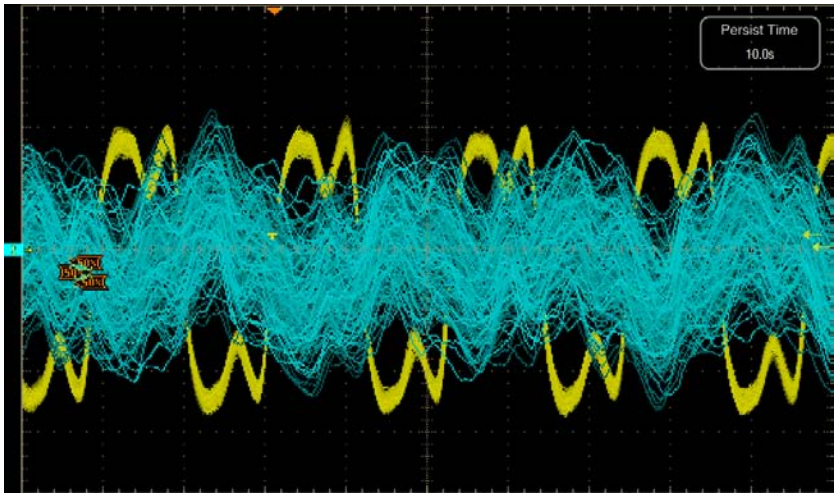
In addition, the TRNG is measuring the difference in phase between multiple oscillators. Even if full injection locking is not taking place, it is merely necessary to overprint the natural oscillator frequencies with an injected frequency. If all the oscillators are running in synchrony, no matter how jittery that synchrony happens to be, any jitter measurement circuit will find no jitter because its time reference also contains the same jitter.

I analyse the attack on relatively prime rings but I demonstrate my attack in 'black box' experiments with no knowledge of device construction. Therefore I assume that I am reducing jitter by the effect of equalising jitter between rings, rather than a reduction of jitter in the whole system.

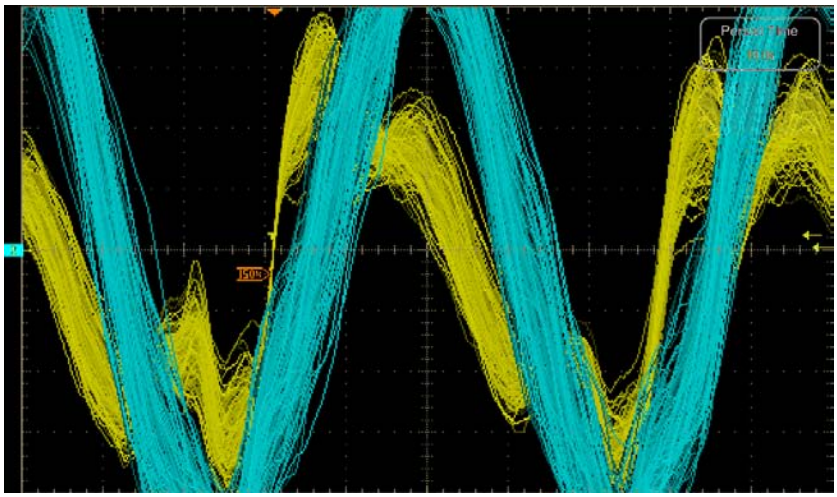
Yoo, Karakoyunlu, Birand, and Sunar (2008) describes locking effects due to poor layout, but generally not in an adversarial manner. They investigate changing the DC supply voltage, but not its AC components. Sunar, Martin, and Stinson (2007) considers active glitch attacks, and concludes these are only able to attack a finite number of bits due to their limited duration. The frequency injection attack is much more powerful, since it can attack all bits simultaneously for as long as desired.

6.4 DISCRETE LOGIC MEASUREMENTS

I set out to measure phase differences in two relatively prime rings. Given their primality, the ring outputs should drift past each other, based on a combination of cumulative jitter and the underlying ring frequency-



(a) No injected signal, rings slide past each other



(b) Strong injection, traces lock phase

Figure 6.4: *Injection locking effect shown on oscilloscope persistent display (discrete inverter experiment from Section 6.4 on the preceding page). View dimensions 8 V × 200 ns; 5 V power supply. Triggering on 3-element ring, with 5-element-ring trace in front. Note in particular how resonances are set up in the ring oscillators that increase the amplitude above the power rails from 5 Vp-p to 10 Vp-p.*

differences. For non-locked rings, I expect phase lag to be uniformly distributed. When locked, phase lag will be concentrated on one value.

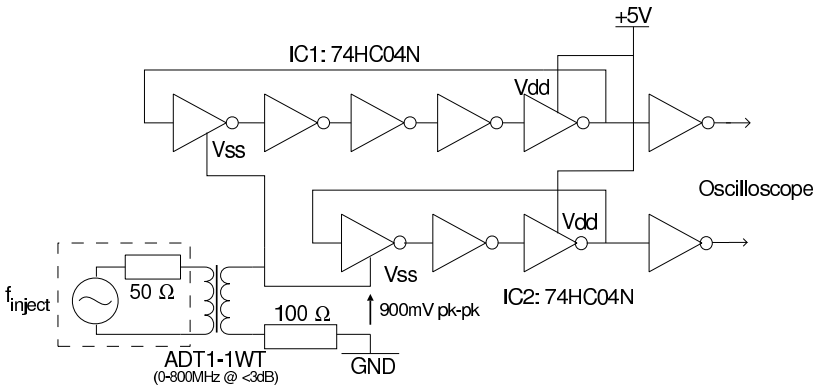
Injection locking is very difficult to simulate in a transient analysis tool such as SPICE (Lai and Roychowdhury 2005). It requires very small timesteps for each oscillation cycle, and a high Q oscillator may require thousands of cycles to lock. When close to the natural frequency, the beat frequency may be a few tens of Hertz. To measure this accurately in a simulation with picosecond-scale steps requires an infeasibly long simulation time. In addition, the asymmetries of a real system will not be modelled in a simulated ideal design.

Due to I/O buffering, it is difficult to measure the behaviour of such fast analogue signals inside an FPGA or ASIC. I first measured the effect in discrete logic. With limited complexity possible, I investigated the simplest ring oscillators: the outputs from three- and five- element rings, with and without frequency injection in the power supply. I used the 74HC04 inverter chip to construct the two mutually-prime rings seen in Figure 6.5a on the next page. Phase lag was measured by triggering an oscilloscope on the rising edge of the three-element ring and measuring the time up to the rising edge of the five-element ring. Such short rings are used in real TRNGs – though they may have a greater region of linear operation than longer rings.

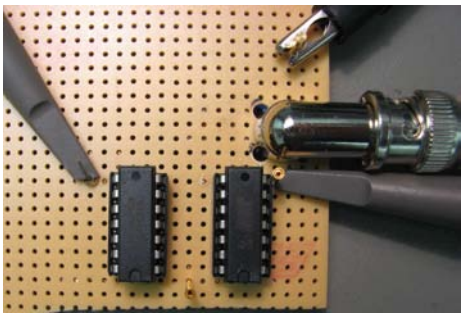
I set up a Tektronix AFG3252 function generator to inject a sine wave at 900 mV pk–pk into the 5 V power rails, and by sweeping frequency I observed locking at 24 MHz. A Tektronix TDS7254B oscilloscope measured the phase lag between the two rings when injecting, and the resulting histograms are plotted in Figure 6.6 on page 176. A very clear clustering around 10 ns can be seen, indicating a lock. This effect is visible in the time domain traces seen in Figure 6.4 on the previous page, which show a marked reduction in the variability of the 5-element output. The slight clustering without injection seen in Figure 6.6 on page 176 is, I believe, due to slightly non-uniform oscilloscope triggering.

6.5 SECURE MICROCONTROLLER

I tested an 8051-compatible secure microcontroller, which has been used in ATMs and other security products. It includes features such as an anti-probing coating and tamper detection, and, at release, claimed to be the most secure product on the market. My example had a date code of 1995, but the device is still recommended for new payment applications by the manufacturer.



(a) Schematic



(b) Test board

Figure 6.5: *Measurement experiment using 74HC04 inverter ICs.*

It provides a hardware TRNG based on frequency differences between two ring oscillators and timing from the user's crystal (11.059 MHz here), and produces 8 bits every 160 µs. 64 bits from the TRNG may be used as the internal encryption key. No further operation details are documented.

I programmed the device to deliver the random bitstream as hexadecimal digits through the serial port, and I displayed it in realtime as a two dimensional black and white image. I adjusted the function generator to inject a sinusoid at 500 mV peak-peak into the chip's power supply as shown in Figure 6.7 on page 177.

By sweeping the frequency, I spotted changes in the patterns produced by the TRNG. The most interesting f_{inject} was at about 1.8 MHz. Obviously-periodic sequences were visible: see Figure 6.8a on page 178–6.8c. In particular the sequence length of the TRNG was controlled by the injected frequency. With another choice of f_{inject} I could also prevent the TRNG returning any values. At no time during any of these tests did the microcontroller otherwise misbehave or detect a fault condition. The

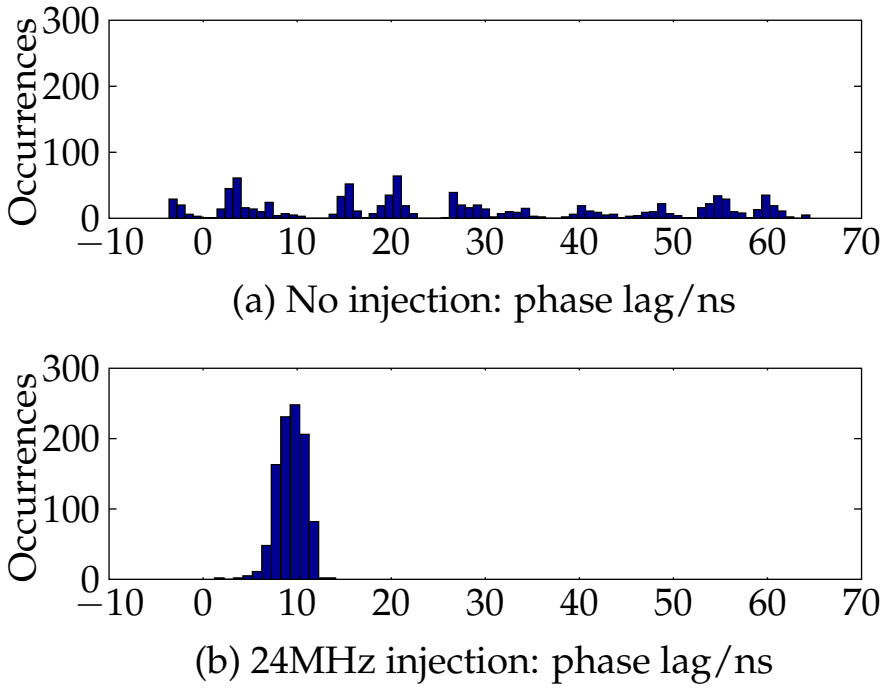


Figure 6.6: *Phase delay between 74HC04 3- and 5- element rings. (a) with no injection, (b) with 24 MHz at 900 mV pk-pk injected into power supply. (25000 samples).*

device is designed to operate at 5 V with a minimum operating voltage of 4.25 V, so it is running within specification.

Figure 6.8b on page 178 indicates a distinct 15-bit long texture, on top of a longer 70-bit sequence. Uncertainty is only present at the overlap between these textures. In a 420-bit sample, I estimate 18 bits of noise. In a 15-bit sequence, that means 0.65 bits may be flipped. The attacker knows the base key is the sequence 010101010..., but not where the 15-bit sequence starts (15 possibilities) or the noise. In most cases, noise is constrained to 3 bits at the start of the 15 bit sequence. In the full 64-bit sequence, the bit flips are $0.65 \times 4 = 2.6$. 3 bits flipped over the whole 64-bit sequence in one of 12 positions gives $\binom{12}{3} = 220$ combinations. Thus I estimate that the total keyspace is smaller than $220 \times 15 = 3300$. In a key length of 32 bits, there are 1.3 bits of noise; the equivalent calculation with 2 bits gives a keyspace of less than $\binom{6}{2} \times 15 = 225 \approx 2^8$.

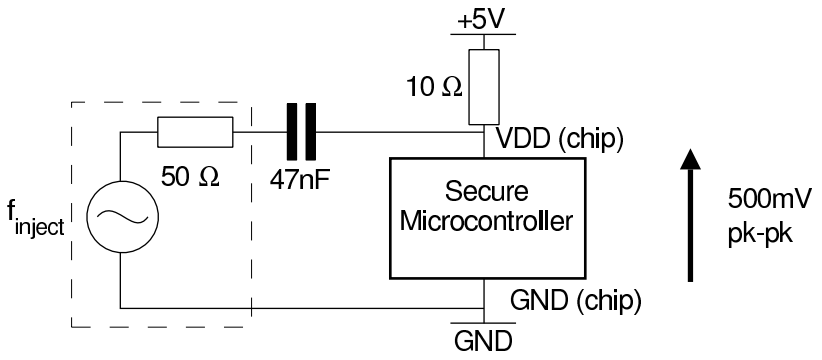


Figure 6.7: *Frequency injection to secure microcontroller.*

6.6 EMV SMARTCARD ATTACK

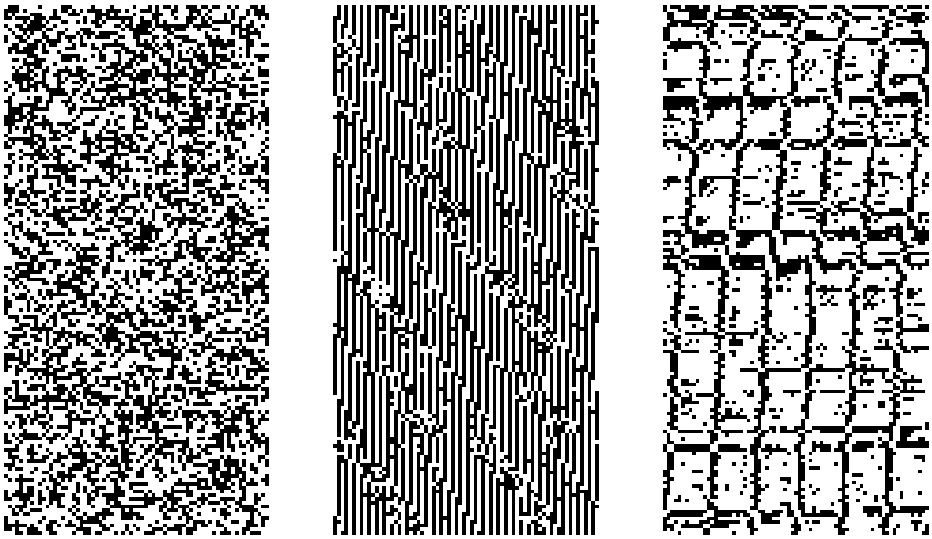
I applied this methodology to an EMV payment card issued in 2004 by a British High Street bank. I picked the first available; with no knowledge of the internals I treated it as a 'black box'. This is a non-invasive attack where no modifications are required to the card under attack.

First I needed to deduce the operating frequency of the card's TRNG. I assumed that such a card would have power analysis protection, so I performed an electromagnetic assessment. An electric field antenna was constructed on a test card. Copper foil was attached beneath the chip as shown in Figure 6.9 on page 179 and Figure 6.10 on page 180, with foil traces between the backside foil patch and the ground pad of the chip. The card was inserted into a Chipdrive Micro 100 card reader connected to a PC via RS232, and standard ISO7816-4 `GET CHALLENGE` commands were used to read the RNG.

I measured three different spectra: (a) not powered or attached to reader (background interference); (b) powered, attached to reader but held in reset and not clocked; and (c) when reading random numbers.

Since a ring oscillator is likely to remain powered even when the card is not initialised, I looked for frequencies that existed when the card was inserted into the reader and unclocked, but not present when the card was removed. I found four such frequencies in the range 0–500 MHz; I chose f_{inject} to be 24.04 MHz, the only frequency below 100 MHz. As this is a black-box system, I do not know if this frequency is optimal; it is merely the first frequency I tried.

I modified the reader to inject a signal as shown in Figure 6.11 on page 180, and I observed the random number statistics. Sweeping the injected frequency and graphically viewing random bits, I saw no obvious



(a) No injection, 70 × 140 bits

(b) 1.822880 MHz injection, 70 × 140 bits

(c) 1.929629 MHz injection, 73 × 146 bits

Figure 6.8: *TRNG bitstream from secure microcontroller with frequency injection, raster scanning left-to-right then top-to-bottom. Bit-widths chosen to illustrate sequences found. Recording into SRAM of 28KB of sequential random bytes at maximum rate, later replayed through the serial port as a hexadecimal string.*

pattern changes. However, statistical analysis of the random data revealed that injecting f_{inject} at 1 V pk-pk across the card powered at 5 V caused the random function to skew. At all times during measurement, the card continued to respond correctly to ISO7816-4 commands and would perform EMV transactions while under attack.

The statistics were analysed using all the tests in the NIST (Rukhin, Soto, Nechvatal, Smid, Barker, Leigh, Levenson, Vangel, Banks, Heckert, Dray, and Vo 2008) and Dieharder (Brown and Eddelbuettel undated)[2] test suites using 1.56×10^9 bits. An outline of both sets of results are shown in Table 6.1 on the next page, with the full tabulated NIST results in Table 6.2 on page 181. By failing most of the tests, I can see that the sequence has become non-random. The FFT test reveals periodicities of around 2000 and 15000 bits. The Rank test, where a 32×32 matrix of random bits should have a rank > 28 (true for my control data), fails with many ranks as low as 19, implying rows or columns are not independent.

[2]Dieharder version 2.28.1, a superset of the DIEHARD suite

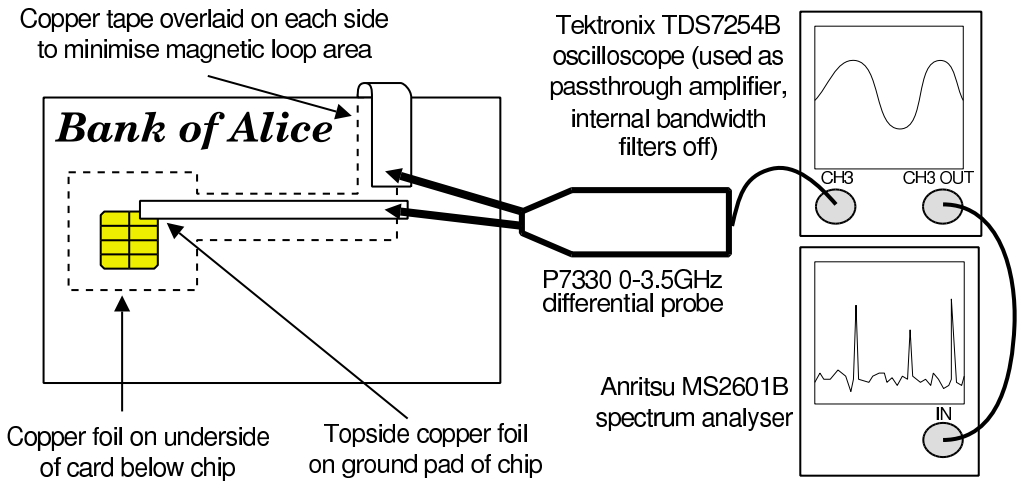


Figure 6.9: *Electric field characterisation of EMV smartcard*

Table 6.1: *Statistical test results from injection into EMV card*

NIST	**Pass**			**Fail**
No injection	187			1
Injection	28			160
Dieharder	**Pass**	**Poor**	**Weak**	**Fail**
No injection	86	6	6	9
Injection	28	16	5	58

6.7 RECOMMENDATIONS AND FURTHER WORK

6.7.1 *Optimisation of the attack*

In the Introduction I outlined an attack on the EMV payment system, which works whether the smartcard uses Static or Dynamic Data Authentication protocols.

An ATM is a computer in a safe, comprising either a PC or custom circuit boards. At least one ATM uses the secure microcontroller I tested as its cryptoprocessor. ATM physical security focuses on preventing access to the money held inside, but this attack needs no access to the cash compartment. Adding injection hardware involves adding a tap to one wire on the PCB – this could be done by an insider or simply by picking the mechanical locks. June 2009 reports (Mills 2009) uncovered malware in ATMs installed by insiders, while in another case (Bogdanich 2003) an at-

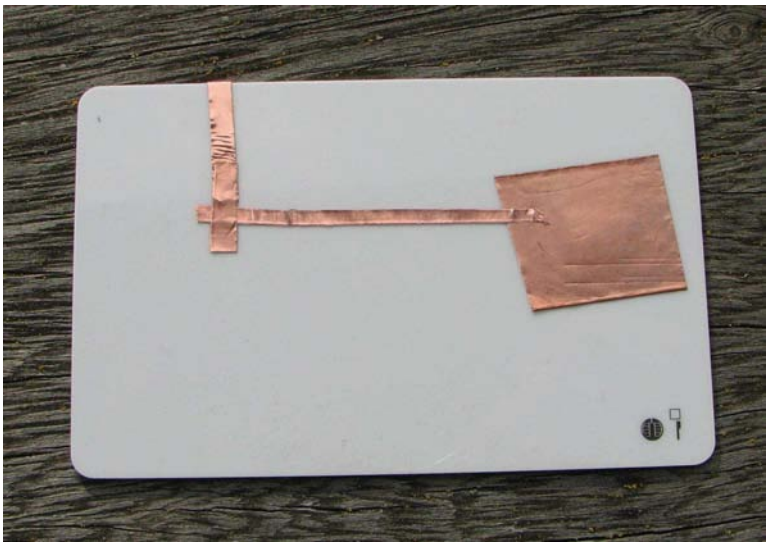


Figure 6.10: *Electric field antenna on underside of example smartcard*

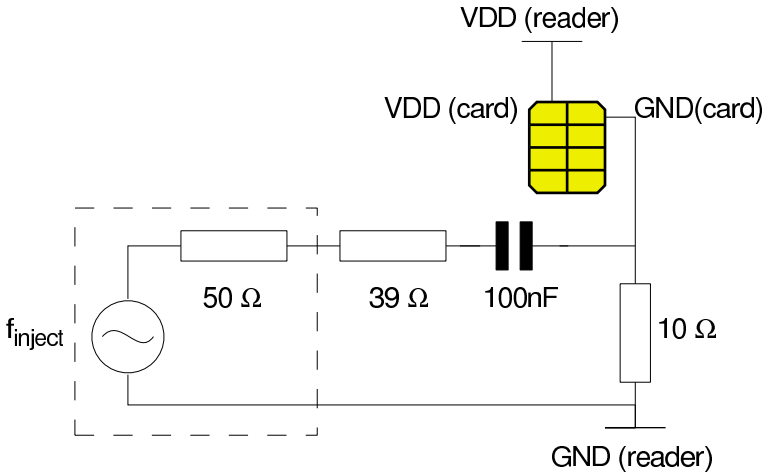


Figure 6.11: *Smartcard frequency injection circuit*

tacker bought up 'white-label' ATMs (normally found in shops and bars), fitted internal recording devices and resold them.

The required number of transactions is small and therefore unlikely to raise alerts at the bank, which is afraid of false alarms. Customers complain about false positives, so there is commercial pressure to be lenient. If a cash withdrawal is performed before the card is used again by the customer, the bank has no way of knowing the transaction was recorded earlier. ATMs are typically only serviced when they go wrong. Even if my

Table 6.2: *NIST results from EMV smartcard.*

	No injection			**Apply f_{inject}**		
	χ^2 P-value	Passes	Result	χ^2 P-value	Passes	Result
Frequency	0.3215	97.44%	PASS	0.0000	21.54%	FAIL
Block Frequency	0.6262	98.97%	PASS	0.0000	0.51%	FAIL
Cumulative Sums	0.2904	97.95%	PASS	0.0000	22.05%	FAIL
Cumulative Sums	0.3902	97.95%	PASS	0.0000	21.54%	FAIL
Runs	0.3811	99.49%	PASS	0.0000	40.00%	FAIL
Longest Run	0.3548	98.97%	PASS	0.0000	73.85%	FAIL
Rank	0.5501	100.00%	PASS	0.0000	0.00%	FAIL
FFT	0.0001	100.00%	PASS	0.0000	0.51%	FAIL
Non-Overlapping Template[a]	0.4523	99.00%	PASS	0.0000	90.89%	FAIL
Overlapping Template	0.4470	98.97%	PASS	0.0000	9.23%	FAIL
Universal	0.0211	98.97%	PASS	0.1488	98.46%	PASS
Approximate Entropy	0.1879	98.97%	PASS	0.0000	1.54%	FAIL
Random Excursions[b]	0.3050	99.26%	PASS	0.2836	99.50%	PASS
Random Excursions Variant[c]	0.4922	99.39%	PASS	0.3053	99.56%	PASS
Serial	0.1168	100.00%	PASS	0.0000	0.00%	FAIL
Serial	0.5501	98.97%	PASS	0.0000	0.00%	FAIL
Linear Complexity	0.0358	99.49%	PASS	0.9554	98.46%	PASS

[a] Mean over 148 tests
[b] Mean over 8 tests
[c] Mean over 18 tests

Dataset of 195×10^6 random bytes. NIST performed 195 runs, each using fresh 10^6 bytes. Minimum pass rate: 96.86% except Random Excursions (96.25% no injection, 93.03% injected).

proposed frequency injector could be spotted by a technician, it may be many months before they are on site.

While I developed my attack with laboratory equipment, the cost of characterising each smartcard, card reader or ATM processor can be made very low. An electric field antenna may be fitted inside a commercial reader, so that nothing is fixed to the card surface. A commercial tunable radio receiver may be attached to the antenna to scan for frequencies of interest, while the frequency synthesiser in a similar receiver may be modified as a cheap way to generate injection frequencies. Given a quantity of identical cards (cheaply acquired on the black market, having little value

once expired or cancelled by the bank), the search is easy to parallelise.

Attacking the TRNG on a card can be optimised by listening to other commands it performs. Each card provides an Answer To Reset (ATR): a response from the card software which can also be used to fingerprint its manufacturer (Rousseau undated). I found cards with the same ATR emitted the same frequencies, most likely if they were built on the same operating system/hardware combination. After characterisation, the attacker can decide which frequencies to inject to a live card based on the ATR. This logic can be built into a terminal or an ATM tap; interception of the card serial data line will reveal the ATR.

Due to electromagnetic interference (EMI) regulations, devices are designed to operate in the presence of interference. Neither of the commercial devices tested failed to operate at any point during these attacks. It is difficult therefore to see how TRNGs could actively detect such attacks without compromising their EMI immunity.

6.7.2 Defences

I have demonstrated that this attack allows the keyspace to be reduced to a size that can be easily brute-forced. As soon as the attacker knows some plaintext, the key may be easily found. The simplest defence is to prevent a brute-force attack. Therefore the best system allows few permitted guesses, which increases the risks for the attacker. Preventing the attacker gaining large quantities of random material would also prevent device characterisation.

To prevent interference a device can filter injected frequencies. Voltage regulation, or merely extra power supply smoothing, may prevent locking, or shielding may be required for electromagnetic attacks. This may be costly and bulky. Devices could refuse to operate at their known-vulnerable frequencies. While this may render them less EMI-immune, it may be acceptable in high security applications. TRNG designs where the feedback loop is combined with logical or register elements (Sunar 2009) may be sufficient to break the locking effect.

Designers can work towards preventing locking by reducing asymmetries in the rings. Carefully balanced transistors may be used, as may equal tapping points on each node. Also, the differential ring oscillator is less affected by supply and substrate noise (Herzel and Razavi 1999). It may be feasible to use this instead of the single-ended ring commonly used. Careful design is required to ensure that reducing susceptibility does not destroy the source of entropy. Hajimiri, Limotyrakis, and Lee (1999) in-

dicates the differential oscillator increases the phase noise, which may be beneficial.

6.7.3 Further work

I have outlined the principles of this attack, but there are many ways in which it could be refined.

The microcontroller used in this attack was chosen as it was rumoured to have a naïve TRNG implementation. The smartcard I tried was not selected at all, it was just the first one to hand. I did not know anything about the TRNG construction at all but that the card was manufactured in 2003, 8 years after the microcontroller. Either this choice was extremely lucky, or the technique has some application to arbitrary TRNGs where the construction method is not known. While the theory behind the technique is easiest to analyse on simple rings as I have done, and it appears that devices using similar construction might also be vulnerable, it is not clear how it affects more involved RNG schemes with latching and gating in the feedback loop such as those described in the literature. Other analogue RNG schemes, such as amplification of noise from semiconductor diodes, while not affected by this oscillator-specific attack, might be similarly susceptible to overriding/jamming the noise signal when exposed to external fields. Given the difficulties of simulation, measurements on physical systems would most likely be required.

Further analysis of the effect of power supply injection is necessary. In the literature, injection locking has mostly been analysed through direct coupling to the signal undergoing oscillation; here, I have used a different mode of coupling, by co-ordinated biasing of the gates it passes through. It would be instructive to determine the minimum power required for this attack and, in particular, how much it can be reduced by on-chip filtering. There are some well-known defences against passive power analysis; it would be interesting to evaluate these for protecting against frequency injection attacks.

It may also be feasible to perform this attack via high-powered electromagnetic radiation, which is more easily focused, and more difficult to mitigate than a conducted attack into the power supply. This could be done by using magnetic loops to induce currents into the device at the appropriate frequencies, or by using the device itself to demodulate the injected frequency (such as a 3 GHz carrier amplitude-modulated by 1.8 MHz); the carrier will more readily propagate, but may be filtered away by parasitic capacitance on the chip leaving the 1.8 MHz harmonic.

The microwave carrier will easily pass through ventilation slots, allowing a non-invasive attack.

Systems with identical ring lengths may be particularly vulnerable due to their shared resonant frequencies. There is further scope for directing this attack if the ring geometries are known. Figure 6.8 on page 178 shows some texture of my TRNG; it may be interesting to use this approach to reverse engineer a TRNG's design from the bitstream.

6.8 SUMMARY

In this chapter I have outlined an attack on ring-oscillator based random number generators using frequency injection I have described the effect and measured its consequences on a security microcontroller used in the EMV system, and in an EMV card. I believe this is an important effect for which all designers of random number generators should test.

Chapter 7
Conclusions

The focus of this research has been the analysis and exploitation of side-channel attacks, with a particular emphasis on electromagnetic analysis (EMA). The research questions concerned how EMA may best be measured, how its behaviour may be characterised, and whether there were any new avenues where EMA may be used as an active attack. I was particularly interested in cryptography used for banking applications.

In reviewing the literature I found a large body of work on power analysis, particularly Differential Power Analysis (DPA). DPA's statistical techniques to extract meaningful cryptographic keys from noise are largely orthogonal to the practicalities of performing a side-channel attack, either by measuring the power or EM emanations of a secure device.

I also read work on a large number of side-channel modes, and of active or invasive attacks. In the electromagnetic arena I came across allusions to NONSTOP and HIJACK vulnerabilities, which remain classified by the US military but are believed to be re-radiation of secret signals by nearby transmitters.

To permit EM experiments I constructed a testing laboratory. I was able to measure EM emissions in the time domain and in the frequency domain. Due to lack of an RF anechoic chamber, I derived techniques for performing EM measurements in the presence of interference. I also constructed a three-axis positioning system to map the EM field in three dimensions. Due to the shared nature of available equipment and evolving experimental setup, I was not able to perform all measurements in identical environments which complicates comparisons.

I designed part of a test chip (the Lochside chip) to aid in EM characterisation. Due to the very tight timescales of the fabrication schedule, I had little time to optimise the antenna test structures I chose for this chip and in the event most were not useful for the experiments I conducted. However I ended up using the clock generators for sensor characterisation, despite their lack of flexibility (in part due to design requirements to reduce spurious harmonics). I developed workarounds for this lack of flexibility as I best I could.

I also used a secure microcontroller (the Springbank chip) from a previous research project. For this chip the designs were fully known and many of the security weaknesses had already been discovered.

7.1 SENSOR EVALUATION

I evaluated a number of sensors for both the electric and magnetic fields. In the electric field I tried a dipole and a foil sheet, the latter giving promising results especially at higher frequencies. In the magnetic field I tried a coil, a commercial anisotropic magnetoresistive (AMR) sensor and various types of hard drive heads. The heads used a number of technologies including inductive and giant magnetoresistance (GMR). Most successful was an inductive hard drive head.

I attempted to calibrate these two sensors using both Lochside and Springbank, and performed simple SPA attacks with them. My justification was that a DPA attack is a superset of an SPA attack, so any sensor showing SPA (even with averaging) is likely useful for DPA. On the other hand, DPA is time consuming to perform for every possible sensor candidate in every configuration. So SPA is a crude analogue for usefulness with DPA.

I also performed three-dimensional field mapping using the inductive hard drive head. Given the choice of test subject, the Lochside chip, I was able to map effects from 'infrastructure' (power, clock) but not from a small on-chip oscillator. Had more time been available I would have progressed to an FPGA which is a more promising target (as has subsequently been demonstrated by Sauvage, Guilley, and Mathieu (2009)).

With hindsight, the evaluation of sensors could have been improved in a number of ways. In particular, since it was undertaken over an extended period of time around other work, the experimental equipment and laboratory conditions had to change over the duration of the experiments. The time to configure the test environment, mechanical mountings, and EM sources was substantial each time the experiments were resumed. The experimental procedure would have been improved if I had further modularised the sensors, so that it was simple and fast to exchange one for another without damage or disturbing the experimental conditions. Then I would have been able to set up the same experiment and evaluate it for multiple sensors to achieve comparable results. Given the fragile nature of many of the sensors, some work on robust mechanical storage and connection would have been necessary to enable easy exchange.

Greater use of commercial sensors would have been beneficial to provide a baseline against which others could be compared. In particular commercial sensors allow easier comparison with other published work. I would also have considered more simple coils, since these are easier to analyse theoretically and predict bounds on how much signal might be expected. Application of the framework of Kuhn (2005) would have been

useful for theoretical analysis.

Having established 3D field mapping apparatus, had more time been available it would have been beneficial to test more than one chip. Indeed the Lochside is not an obvious choice for field mapping given its relatively quiescent state.

7.2 ACTIVE ATTACKS

The field of active attacks was then considered. I conceived the *re-emission attack*, namely the injection of frequencies into a device in the hope of evoking intermodulation. This would enable an attacker to shift up or down the injection frequency to suit their receiver (such as an off-the-shelf commercial receiver). It also enables choice of an optimal frequency for propagation and antenna design, such as a more covert antenna. As recognised by the US military, the accidental effect can be a significant EM threat. The deliberate use of re-emission has thus even more potential efficacy.

I demonstrated that the re-emission attack was possible, both with proof-of-concept attacks on oscillators and the Springbank secure microcontroller. Co-incidentally Burnside, Erdogan, and Arslan (2008) published similar results which demonstrated the validity of this attack. As Burnside was more advanced in his experiments, I decided to use the techniques on a different target.

I conceived a new active attack, the frequency injection attack on True Random Number Generators (TRNGs). This uses the same equipment as the re-emission attack, but aims at disrupting functionality rather than radiating. The principle of injection locking is well known to the oscillator design community. I reasoned that it might be possible to force a number of oscillators to injection lock by application of an external signal into the power supply. In some TRNGs free-running ring oscillators are used as an entropy source, so injection locking reduces the entropy available which is a security risk.

First I successfully achieved the effect on a pair of rings formed from discrete (74HC04) logic gates. Success at altering the ring frequencies demonstrated the potential of the attack, but is far from a good match for the conditions found on an IC.

I then tried a secure 8051-based microcontroller, which is known to be used in banking automatic telling machines (ATMs) and is still recommended for those applications. The effect was very successful, reducing the entropy of a 32-bit unpredicatable number as emitted by the ATM from

approximately 4 billion (2^{32}) to roughly 225. I then described how the ATM cash withdrawal protocol might be used to reduce the average number of tries required to steal money to be 15 ($\sqrt{225}$).

I also tested an EMV banking card. This was simply the first card that was tested – it was not specially selected. I knew nothing about the construction of the card, or that it even used the TRNG design I was targeting. Without injection, the card passed 99.5% of the NIST test suite. With power supply injection it passed 14.9% of tests. The cause of failure was particularly picked up by the Fast Fourier Transform test, which found periodicities, and the Rank test, which found non-independence of nearby data bits.

In theory, it should be possible to find a group of frequencies that hits the eigenfrequencies of all the internal ring oscillators. It would be interesting to investigate whether multiple frequencies are able to have a greater effect than a single harmonic.

In the ATM attack, I outlined a number of ways in which it could be optimised and a number of means of defence. In particular, techniques related to re-emission (such as illuminating with a mix of frequencies to aid propagation, then using the secure device as a parasitic demodulator) may come in useful. It would be interesting to verify these techniques. Also, traditional DPA defences may be useful – further works is required to evaluate them.

In summary, I have conceived a new field of *active electromagnetic attack* to modern cryptographic devices. This field has great potential for the attacker.

BIBLIOGRAPHY

ADLER, R. (1946): "A study of locking phenomena in oscillators", *Proceedings of the IRE and Waves and Electrons*, 34, 351–357.

AGRAWAL, D., ARCHAMBEAULT, B., RAO, J. R., AND ROHATGI, P. (2002a): "The EM Side-Channel(s)", in *Cryptographic Hardware and Embedded Systems - CHES 2002, 4th International Workshop, Redwood Shores, CA, USA, August 13-15, 2002, Revised Papers*, ed. by B. S. Kaliski, Jr., Çetin Kaya Koç, and C. Paar, vol. 2523 of *Lecture Notes in Computer Science*, pp. 29–45. Springer.

——— (2002b): "The EM Side-Channel(s): Attacks and Assessment Methodologies", *IBM Technical Report*.

AHN, Y.-S. (1999): "US Patent 5910861: Technique for controlling the write currents of a magnetic disk recording apparatus".

ALLAN, R. A. (2001): *A history of the personal computer*. Allan Publishing, London, Ontario, Canada.

ANDERSON, R. J. (2001): *Security engineering: a guide to building dependable distributed systems*. John Wiley and Sons, first edn.

ANDERSON, R. J., AND KUHN, M. G. (1996): "Tamper Resistance - a Cautionary Note", in *The Second USENIX Workship on Electronic Commerce Proceedings, Oakland, California*, pp. 1–11. USENIX Association.

——— (1999): "Soft Tempest – An Opportunity for NATO", in *Information Systems Technology (IST) Symposium "Protecting NATO Information Systems in the 21st Century"*, Washington, DC, USA. NATO Research and Technology Organization (RTO).

ASO, K., SATO, T., AND ISHIBASHI, M. (1999): "Magnetic force microscopic study of magnetic tapes recorded at MHz frequencies", *Journal of Magnetism and Magnetic Materials*, 193, 430–433.

ATMEL CORPORATION (2004): "2-Channel 3.3V or 5V GMR Head Preamplifier: AT78C6002", `http://www.atmel.com/dyn/resources/prod_documents/doc3460.pdf`.

BAIBICH, M. N., BROTO, J. M., FERT, A., VAN DAU, F. N., PETROFF, F., ETIENNE, P., CREUZET, G., FRIEDERICH, A., AND CHAZELAS, J. (1988): "Giant Magnetoresistance of (001)Fe/(001)Cr Magnetic Superlattices", *Physical Review Letters*, 61(21), 2472–2475.

BAK, P. (1986): "The Devil's Staircase", *Physics Today*, 39(12), 38–45.

BALLANTINE, S. (1923): *Radio Telephony for Amateurs*. Chapman & Hall Ltd, London, first edn.

BARTLETT, A. C. (1933): "The calculation of modulation products", *Philosophical Magazine Series 7*, 16(107), 845–847.

——— (1934): "The calculation of modulation products – II", *Philosophical Magazine Series 7*, 17(113), 628–633.

BELLARE, M., GOLDWASSER, S., AND MICCIANCIO, D. (1997): ""Pseudo-Random" Number Generation Within Cryptographic Algorithms: The DSS Case", in *CRYPTO*, ed. by B. S. Kaliski, Jr., vol. 1294 of *Lecture Notes in Computer Science*, pp. 277–291. Springer.

BELLO, L. (2008): "DSA-1571-1 openssl – predictable random number generator", *Debian Security Advisory*, `http://www.debian.org/security/2008/dsa-1571`.

BOAK, D. G. (1973): *A History of U.S. Communications Security: The David G. Boak Lectures*, vol. I. National Security Agency, Fort George G. Meade, Maryland, USA.

BOBBETT, D. G. (ed.) (1999): *World Radio TV Handbook 1999*, vol. 53. WRTH Publications, Milton Keynes, UK.

BOGDANICH, W. (2003): "STEALING THE CODE: Con Men and Cash Machines; Criminals Focus on A.T.M.'s, Weak Link in Banking System", *The New York Times*, `http://query.nytimes.com/gst/fullpage.html?res=9803E6DD103EF930A3575BC0A9659C8B63`.

BOND, M. K. (2000): "A Chosen Key Difference Attack on Control Vectors", `http://www.cl.cam.ac.uk/~mkb23/research/CVDif.pdf`.

BONEH, D., DEMILLO, R. A., AND LIPTON, R. J. (2001): "On the Importance of Eliminating Errors in Cryptographic Computations", *Journal of Cryptology*, 14(2), 101–119.

BÖTTCHER, C. (1973): *Theory of Electric Polarization. Volume 1: Dielectrics in static fields.* Elsevier, Amsterdam, second edn.

BRIDGES, G. E., NORUTTUN, D., SAID, R. A., THOMSON, D. J., LAM, T., AND QI, R. (1998): "Non-contact probing of high speed microelectronics using electrostatic force sampling", *Journal of Vacuum Science and Technology*, 16(2), 830–833.

BROWN, M. (2010a): "mb21 – The Transmission Gallery. Cambridge (Madingley)", `http://tx.mb21.co.uk/gallery/madingley.php`.

——— (2010b): "mb21 – The Transmission Gallery. Peterborough. Coverage Area Maps", `http://tx.mb21.co.uk/gallery/peterborough/maps.php`.

BROWN, R. G., AND EDDELBUETTEL, D. (undated): "Dieharder: A Random Number Test Suite", `http://www.phy.duke.edu/~rgb/General/dieharder.php`.

BUCK, F. J. (1772): *Mathematischer Beweiß: daß die Algebra zur Entdeckung einiger verborgener Schriften bequem angewendet werden könne*. Published by the widow of J. D. Zeisen and the heirs of J. H. Hartung, Königsberg, `http://www-math.uni-paderborn.de/~aggathen/Publications/buc72.pdf`.

BULL WORLDWIDE INFORMATION SYSTEMS (1996): "Bull CP8: It's a Smart Card world", `http://web.archive.org/web/19961221161632/http://www.cp8.bull.net/`.

BURNSIDE, A., ERDOGAN, A., AND ARSLAN, T. (2008): "The Re-emission Side Channel", in *Bio-inspired Learning and Intelligent Systems for Security, 2008. BLISS '08. ECSIS Symposium on*, pp. 154–159, Edinburgh, UK.

CAIN, W., PAYNE, A., QIU, G., LATEV, D., IMAI, D., HEMPSTEAD, R., MCNEIL, M., AND PHENICIE, C. (1996): "Achieving 1 Gbit/in^2 with inductive recording heads", *IEEE Transactions on Magnetics*, 32(5), 3551–3553.

CARR, J. (1928): "Radio Noises You Can Cure", *Popular Science Monthly*, 112(2), 69–70.

CARR, J. J. (2001): *Secrets of RF Circuit Design*. McGraw-Hill, New York.

CARUSO, M. J., BRATLAND, T., SMITH, C. H., AND SCHNEIDER, R. (1998): "A New Perspective on Magnetic Field Sensing", *Sensors*, 15(12), 34–46.

CHARI, S., RAO, J. R., AND ROHATGI, P. (2002): "Template Attacks", in *Cryptographic Hardware and Embedded Systems - CHES 2002, 4th International Workshop, Redwood Shores, CA, USA, August 13-15, 2002, Revised Papers*, ed. by B. S. Kaliski, Jr., Çetin Kaya Koç, and C. Paar, vol. 2523 of *Lecture Notes in Computer Science*, pp. 13–28. Springer.

——— (2003): "Multi-channel Attacks", in *Cryptographic Hardware and Embedded Systems - CHES 2003, 5th International Workshop, Cologne, Germany, September 8-10, 2003, Proceedings*, ed. by C. D. Walter, Çetin Kaya Koç, and C. Paar, vol. 2779 of *Lecture Notes in Computer Science*, pp. 2–16. Springer.

COUCH, III, L. W. (2000): *Digital and Analog Communication Systems*. Prentice-Hall, sixth edn.

DETHLOFF, J., AND GRÖTTRUP, H. (1972): "US Patent 3641316: Identification System".

DIENY, B., SPERIOSU, V. S., METIN, S., PARKIN, S. S. P., GURNEY, B. A., BAUMGART, P., AND WILHOIT, D. R. (1991): "Magnetotransport properties of magnetically soft spin-valve structures (invited)", *35th Annual Conference on Magnetism and Magnetic Materials*, 69(8), 4774–4779.

DRIMER, S., MURDOCH, S. J., AND ANDERSON, R. (2008): "Thinking inside the box: system-level failures of tamper proofing", Technical Report UCAM-CL-TR-711, University of Cambridge Computer Laboratory.

EASTLAKE, D., SCHILLER, J., AND CROCKER, S. (2005): "Best Common Practice 106: Randomness Requirements for Security", Technical Report BCP106, Internet Engineering Task Force.

ECRYPT NETWORK OF EXCELLENCE (undated): "Side Channel Cryptanalysis Lounge", `http://www.crypto.ruhr-uni-bochum.de/en_sclounge.html`.

ELLINGBOE, J. K. (1972): "US Patent 3637994: Active Electrical Card Device".

EMERY, R. (undated): "UK-Dab.info: What can I receive?", `http://www.uk-dab.info/stationlist.php`.

EMVCO, LLC (2007): "EMV Issuer and Application Security Guidelines", (version 2.1).

——— (2008): "EMV 4.2 Specification", `http://www.emvco.com/`.

FAIRBANKS, S., AND MOORE, S. (2005): "Self-timed circuitry for global clocking", in *Eleventh International Symposium on Advanced Research in Asynchronous Circuits and Systems (ASYNC05)*, pp. 86–96, New York City, USA. IEEE Computer Society Press.

FOURNIER, J. J. A., MOORE, S. W., LI, H., MULLINS, R. D., AND TAYLOR, G. S. (2003): "Security Evaluation of Asynchronous Circuits", in *Cryptographic Hardware and Embedded Systems - CHES 2003, 5th International Workshop, Cologne, Germany, September 8-10, 2003, Proceedings*, ed. by C. D. Walter, Çetin Kaya Koç, and C. Paar, vol. 2779 of *Lecture Notes in Computer Science*, pp. 137–151. Springer.

G3CARD CONSORTIUM (2003): "G3 Card: Public Final Report", Technical Report IST-1999-13515, EU Information Society Technologies.

GANDOLFI, K., MOURTEL, C., AND OLIVIER, F. (2001): "Electromagnetic Analysis: Concrete Results", in *Cryptographic Hardware and Embedded Systems - CHES 2001, Third International Workshop, Paris, France, May 14-16, 2001, Proceedings*, ed. by Çetin Kaya Koç, D. Naccache, and C. Paar, vol. 2162 of *Lecture Notes in Computer Science*, pp. 251–261. Springer.

GEBOTYS, C. H., HO, S., AND TIU, C. C. (2005): "EM Analysis of Rijndael and ECC on a Wireless Java-Based PDA", in *Cryptographic Hardware and Embedded Systems - CHES 2005, 7th International Workshop, Edinburgh, UK, August 29 - September 1, 2005, Proceedings*, ed. by J. R. Rao, and B. Sunar, vol. 3659 of *Lecture Notes in Computer Science*, pp. 250–264. Springer.

GIRARDOT, Y. (1984): "Bull CP8 Smart Card Uses in Cryptology", in *Advances in Cryptology: Proceedings of EUROCRYPT 84, A Workshop on the Theory and Application of of Cryptographic Techniques, Paris, France, April 9-11, 1984, Proceedings*, ed. by T. Beth, N. Cot, and I. Ingemarsson, vol. 209 of *Lecture Notes in Computer Science*, p. 464. Springer.

GOVINDAVAJHALA, S., AND APPEL, A. W. (2003): "Using Memory Errors to Attack a Virtual Machine", in *IEEE Symposium on Security and Privacy*, pp. 154–165. IEEE Computer Society.

GREGSON, S., MCCORMICK, J., AND PARINI, C. (2007): *Principles of Planar Near-Field Antenna Measurements*. Institution of Engineering and Technology, London, UK.

GUILLOU, L. (2004): "Histoire de la carte à puce du point de vue d'un cryptologue", in *Actes du Septième Colloque sur l'Histoire de l'Informatique et des Transmissions'*, pp. 126–154, Cesson–Rennes, France.

HAGHIRI, Y., AND TARANTINO, T. (2002): *Smart Card Manufacturing*. John Wiley and Sons, Chichester, UK.

HAJIMIRI, A., LIMOTYRAKIS, S., AND LEE, T. H. (1999): "Jitter and Phase Noise in Ring Oscillators", *IEEE Journal of Solid-State Circuits*, 34(6), 790–804.

HALPERN, J. W. (1975): "US Patent 3906460: Proximity data transfer system with tamper proof portable data token".

HARTMANN, U. (1999): "Magnetic Force Microscopy", *Annual Review of Materials Science*, 29, 53–87.

HENDRY, M. (2001): *Smart Card Security and Applications*. Artech House, Norwood, MA, USA, second edn.

HERZEL, F., AND RAZAVI, B. (1999): "A Study of Oscillator Jitter Due to Supply and Substrate Noise", *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 46(1), 56–42.

HESS, F. M., ET AL. (undated): "Linux GPIB Project", `http://linux-gpib.sourceforge.net/`.

HILL, L. S. (1929): "Cryptography in an Algebraic Alphabet", *The American Mathematical Monthly*, 36(6), 306–312.

HONEYWELL INTERNATIONAL INC (2008): "1- and 2-Axis Magnetic Sensors HMC1001/1002/1021/1022".

HSUEH, M.-C., TSAI, T. K., AND IYER, R. K. (1997): "Fault Injection Techniques and Tools", *Computer*, 30(4), 75–82.

HUYGHENS, C. (1673): *Horologium Oscillatorium, siue De Motu Pendulorum ad Horologia Aptato Demostrationes Geometricæ*. F. Muguet, Paris.

IBM STORAGE TECHNOLOGY DIVISION (2000): "IBM Deskstar 75GXP and Deskstar 40GV hard disk drives".

ISO/IEC (2009): "ISO/IEC 15408: Information technology – Security techniques – Evaluation criteria for IT security".

KAHN, D. (1996): *The Codebreakers: The Story of Secret Writing*. Scribner, New York.

KAZIMIERCZUK, M. K. (2009): *High-Frequency Magnetic Components*. John Wiley and Sons.

KELSEY, J., SCHNEIER, B., WAGNER, D., AND HALL, C. (2000): "Side Channel Cryptanalysis of Product Ciphers", *Journal of Computer Security*, 8, 141–158.

KERCKHOFFS, A. (1883): "La cryptographie militaire I", *Journal des Sciences Militaires*, 9, 5–38.

KIRTLEY, J. R., AND JOHN P. WIKSWO, J. (1999): "Scanning SQUID Microscopy", *Annual Reviews of Materials Science*, 29, 117–148.

KOCHER, P. C. (1996): "Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS, and Other Systems", in *Proceedings of the 16th Annual International Cryptology Conference on Advances in Cryptology*, pp. 104–113. Springer.

KOCHER, P. C., JAFFE, J., AND JUN, B. (1999): "Differential Power Analysis", in *Proceedings of the 19th Annual International Cryptology Conference on Advances in Cryptology*, pp. 388–397. Springer.

KRAUS, J. D. (1991): *Electromagnetics*. McGraw-Hill, fourth edn.

KRAUS, J. D., AND MARHEFKA, R. J. (2001): *Antennas for All Applications*. McGraw-Hill, third edn.

KUHN, M. G. (2003): "Compromising emanations: eavesdropping risks of computer displays", Technical Report UCAM-CL-TR-577, University of Cambridge Computer Laboratory.

——— (2005): "Security Limits for Compromising Emanations", in *Cryptographic Hardware and Embedded Systems - CHES 2005, 7th International Workshop, Edinburgh, UK, August 29 - September 1, 2005, Proceedings*, ed. by J. R. Rao, and B. Sunar, vol. 3659 of *Lecture Notes in Computer Science*, pp. 265–279. Springer.

LAI, X., AND ROYCHOWDHURY, J. (2005): "Analytical equations for predicting injection locking in LC and ring oscillators", in *IEEE 2005 Custom Integrated Circuits Conference*, pp. 461–464.

LARK-HOROVITZ, K., AND JOHNSON, V. A. (eds.) (1959): *Methods of Experimental Physics. Volume 6, Part B: Solid State Physics.* Academic Press, New York.

LI, H., MARKETTOS, A. T., AND MOORE, S. W. (2005): "Security Evaluation Against Electromagnetic Analysis at Design Time", in *Cryptographic Hardware and Embedded Systems - CHES 2005, 7th International Workshop, Edinburgh, UK, August 29 - September 1, 2005, Proceedings*, ed. by J. R. Rao, and B. Sunar, vol. 3659 of *Lecture Notes in Computer Science*, pp. 280–292. Springer.

LOUGHRY, J., AND UMPHRESS, D. A. (2002): "Information Leakage from Optical Emanations", *ACM Transactions on Information and System Security*, 5(3), 262–289.

MACDONALD, S. B., AND GASTMANN, A. L. (2004): *A History of Credit and Power in the Western World*. Transaction Publishers, New Brunswick, New Jersey, USA.

MANGARD, S., OSWALD, E., AND POPP, T. (2007): *Power Analysis Attacks: Revealing the Secrets of Smart Cards*. Springer, New York.

MAXWELL, J. C. (1861): "On Physical Lines of Force", *Philsophical Magazine Series 4*, 21(139), 161–175.

MESGARZADEH, B., AND ALVANDPOUR, A. (2005): "A study of injection locking in ring oscillators", *Proceedings of IEEE International Symposium on Circuits and Systems, 2005*, 6, 5465–5468.

MESSERGES, T. S., DABBISH, E. A., AND SLOAN, R. H. (1999): "Investigations of Power Analysis Attacks on Smartcards", in *Proceedings of the USENIX Workshop on Smartcard Technology*, Chicago, Illinois, USA. USENIX Association.

MESSMER, E. (1995): "Credit card firms plan 'digital money' future", *Network World*, 12(17), 47, 103.

MEULSTEE, L. (undated): "Wireless for the Warrior: Fullerphone", `http://wftw.nl/fullerphone/fullerphone.html`.

MILLIGAN, T. A. (2005): *Modern antenna design*. Wiley, New Jersey, second edn.

MILLS, E. (2009): "Hacked ATMs let criminals steal cash, PINs", *ZDNet UK*, http://news.zdnet.co.uk/security/0,1000000189,39660339,00.htm.

MOCK, J., ET AL. (2002): "GPIB module for Perl", http://www.mock.com/gpib/.

MOORE, S., ANDERSON, R., MULLINS, R., TAYLOR, G., AND FOURNIER, J. J. A. (2003): "Balanced Self-Checking Asynchronous Logic for Smart Card Applications", *Microprocessors and Microsystems Journal*, 27(7), 421–430.

MORENO, R. (1975): "French Patent FR2266222: Procédé et dispositif de commande électronique".

MULLINS, R., WEST, A., AND MOORE, S. (2004): "Low-Latency Virtual-Channel Routers for On-Chip Networks", in *ISCA '04: Proceedings of the 31st Annual International Symposium on Computer Architecture*, pp. 188–197, Washington, DC, USA. IEEE Computer Society.

MURDOCH, S. J., DRIMER, S., ANDERSON, R., AND BOND, M. (2010): "Chip and PIN is Broken", in *Proceedings of 2010 IEEE Symposium on Security and Privacy*, pp. 433–446.

NACCACHE, D., AND TUNSTALL, M. (2000): "How to Explain Side-Channel Leakage to Your Kids", in *Cryptographic Hardware and Embedded Systems - CHES 2000, Second International Workshop, Worcester, MA, USA, August 17-18, 2000, Proceedings*, ed. by Çetin Kaya Koç, and C. Paar, vol. 1965 of *Lecture Notes in Computer Science*, pp. 229–230. Springer.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (1996): "Galileo FAQ – Galileo's Antennas", http://www2.jpl.nasa.gov/galileo/faqhga.html.

NATIONAL FREQUENCY PLANNING GROUP (2008): "United Kingdom Frequency Allocation Table", Issue 15.

NATIONAL INSTITUE OF SCIENCE AND TECHNOLOGY (2001): "FIPS 140-2: Security Requirements for Cryptographic Modules".

NATIONAL SECURITY AGENCY (1972): "TEMPEST: A Signal Problem", *Cryptologic Spectrum*, Summer 1972, 26–30.

——— (1975): "NACSEM 5112 NONSTOP Evaluation Techniques".

——— (199x): "NSA/CSS Regulation: NSA/CSS 90-6", year is illegible in original document.

NATIONAL SECURITY TELECOMMUNICATIONS AND INFORMATION SYSTEMS SECURITY (1995): "Red/Black Installation Guidance", `http://cryptome.org/tempest-2-95.htm`.

NEVE, M., PEETERS, E., SAMYDE, D., AND QUISQUATER, J.-J. (2003): "Memories: A Survey of Their Secure Uses in Smart Cards", in *Second International IEEE Security in Storage Workshop (SISW 2003), Information Assurance, The Storage Security Perspective, 31 October 2003, Washington, DC, USA*, pp. 62–72. IEEE Computer Society.

NEWPORT CORPORATION (undated): "Motion Basics and Standards", `http://www.newport.com/Motion-Basics-and-Standards/140230/1033/catalog.aspx`.

NG, H.-T., FARJAD-RAD, R., LEE, M.-J. E., DALLY, W. J., GREER, T., POULTON, J., EDMONDSON, J. H., RATHI, R., AND SENTHINATHAN, R. (2003): "A Second-Order Semidigital Clock Recovery Circuit Based on Injection Locking", *IEEE Journal of Solid-State Circuits*, 38(12), 2101–2110.

PC WORLD BUSINESS (2010): "Eltron P310i plastic card printer", `http://www.pcwb.co.uk/catalogue/item/A0410456`.

POPULAR MECHANICS (1929): "Locating By-Pass Condensers", *Popular Mechanics*, 52(2), 310.

PREMIER FARNELL (2005): "RG58/RG59 Series Coaxial Cables".

QUEENS' COLLEGE, CAMBRIDGE (2000): "Cambridge Radio & TV", `http://www.quns.cam.ac.uk/Queens/CompFacil/radiotv/`.

QUISQUATER, J.-J., AND SAMYDE, D. (2001): "ElectroMagnetic Analysis (EMA): Measures and Counter-Measures for Smart Cards", in *Proceedings of E-smart 2001*, ed. by I. Attali, and T. Jensen, vol. 2140 of *Lecture Notes in Computer Science*, pp. 200–210. Springer-Verlag.

RANKL, W., AND EFFING, W. (2004): *Smart Card Handbook*. John Wiley and Sons, Chichester, UK.

RIVEST, R. L., SHAMIR, A., AND ADLEMAN, L. (1978): "A method for obtaining digital signatures and public-key cryptosystems", *Communications of the ACM*, 21(2), 120–126.

ROBINSON, F. N. H. (1973): *Electromagnetism*. Clarendon Press, Oxford, UK.

ROUSSEAU, L. (undated): "ATR to card type table from pcsc_tools package", `http://ludovic.rousseau.free.fr/softwares/pcsc-tools/smartcard_list.txt`.

RUKHIN, A., SOTO, J., NECHVATAL, J., SMID, M., BARKER, E., LEIGH, S., LEVENSON, M., VANGEL, M., BANKS, D., HECKERT, A., DRAY, J., AND VO, S. (2008): "A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications", Technical Report SP800-22, National Institute of Standards and Technology, USA.

SAMSUNG ELECTRONICS (1997): "WINNER-1 Project WN31601A: New 1.61 Gbyte High Performance 3.5" – 1" height disk drive", `http://www.samsung.com/global/business/hdd/products/downloads/wnr31601a.pdf`.

SAMYDE, D., SKOROBOGATOV, S., ANDERSON, R., AND QUISQUATER, J.-J. (2002): "On a New Way to Read Data from Memory," in *First International IEEE Security in Storage Workshop (SISW 2002), 11 December 2002, Greenbelt, Maryland, USA*, pp. 65–69. IEEE Computer Society.

SAUVAGE, L., GUILLEY, S., AND MATHIEU, Y. (2009): "Electromagnetic Radiations of FPGAs: High Spatial Resolution Cartography and Attack on a Cryptographic Module", *ACM Transactions on Reconfigurable Technology and Systems*, 2(1), 1–24.

SCHMIDT, J.-M. (2008): "Differential Fault Analysis", Technical report, A-SIT: Secure Information Technology Center - Austria.

SCHMIDT, J.-M., AND HUTTER, M. (2007): "Optical and EM Fault-Attacks on CRT-based RSA: Concrete Results", in *Austrochip 2007 - Proceedings of the 15th Austrian Workshop on Microelectronics*, ed. by K. C. Posch, and J. Wolkerstorfer, Graz, Austria. Verlag der Technischen Universität Graz.

SCHNEIER, B. (1996): *Applied Cryptography*. John Wiley and Sons, second edn.

SCIENGINES GMBH (2009): "Break DES in less than a single day", `http://www.sciengines.com/company/news-a-events/74-des-in-1-day.html`.

SHAMIR, A., AND TROMER, E. (undated): "Acoustic cryptanalysis: On nosy people and noisy machines", http://people.csail.mit.edu/tromer/acoustic/.

SHARP MICROELECTRONICS (1999): "LH77790B Embedded Microcontroller User's Guide".

SHARPLES, S. D., ET AL. (2008): "VXI11 Ethernet Protocol for Linux", http://optics.eee.nottingham.ac.uk/vxi11/.

SHIRVANI, P. P. (2001): "Fault-Tolerant Computing for Radiation Environments", Technical Report 01-6, Stanford University, Stanford, CA, USA.

SKOROBOGATOV, S., AND KUHN, M. (2005): "Power analysis of the Motorola MC68HC908AZ60A microcontroller".

SKOROBOGATOV, S. P. (2009): "Using Optical Emission Analysis for Estimating Contribution to Power Analysis", in *Sixth Workshop on Fault Diagnosis and Tolerance in Cryptography (FDTC 2009), 6th September 2009, Lausanne, Switzerland*.

SKOROBOGATOV, S. P., AND ANDERSON, R. J. (2003): "Optical Fault Induction Attacks", in *Cryptographic Hardware and Embedded Systems - CHES 2002, 4th International Workshop, Redwood Shores, CA, USA, August 13-15, 2002, Revised Papers*, ed. by B. S. K. Jr., vol. 2523 of *LNCS2523*, pp. 2–12. Springer-Verlag.

SMITH, D. C. (2003): "Technical Tidbit July 2003: Measuring E-Field Coupled IC Chip Noise", http://emcesd.com/tt2003/tt070103.htm.

SOLTANI, P. K., WYSNEWSKI, D., AND SWARTZ, K. (1999): "Amorphous Selenium Direct Radiography", in *Proceedings of Computerized Tomography for Industrial Applications and Image Processing in Radiology*, pp. 123–132, Berlin, Germany. Deutsche Gesellschaft Für Zerstörungsfreie Prüfung.

SOUTHERN UTAH UNIVERSITY (2004): "SUU In View: The Magazine for Southern Utah University", Fall 2004.

SUNAR, B. (2009): "True Random Number Generators for Cryptography", in *Cryptographic Engineering*, ed. by Ç. K. Koç, chap. 4, pp. 55–74. Springer.

SUNAR, B., MARTIN, W. J., AND STINSON, D. R. (2007): "A Provably Secure True Random Number Generator with Built-In Tolerance to Active Attacks", *IEEE Transactions on Computers*, 56(1), 109–119.

TAKAHASHI, T., AND KIMURA, T. (2000): "US Patent 6141168: Automatic calibration method, read apparatus and storage apparatus".

TEXAS INSTRUMENTS STORAGE PRODUCTS GROUP (1998a): "SSI 32P4910A PRML Read Channel with PR4, 8/9 ENDEC, 4-Burst Servo".

——— (1998b): "SSI 32R2210R/11R/12R +5V, 2-, 4-channel thin film read/write device".

THONG, J. T. (2004): "Electron Beam Probing", in *Microelectronics failure analysis: desk reference*, ed. by EDFAS Desk Reference Committee, chap. 9, pp. 439–444. ASM International, fifth edn.

TRENT, R. L. (1952): "A Transistor Reversible Binary Counter", *Proceedings of the IRE*, 40(11), 1562–1572.

TSUNOO, Y., SAITO, T., SUZAKI, T., SHIGERI, M., AND MIYAUCHI, H. (2003): "Cryptanalysis of DES Implemented on Computers with Cache," in *Cryptographic Hardware and Embedded Systems - CHES 2003, 5th International Workshop, Cologne, Germany, September 8-10, 2003, Proceedings*, ed. by C. D. Walter, Çetin Kaya Koç, and C. Paar, vol. 2779 of *Lecture Notes in Computer Science*, pp. 62–76. Springer.

UGON, M. (1980): "US Patent 4211919: Portable data carrier including a microprocessor".

UNITED STATES AIR FORCE (1998): "Air Force Systems Security Memorandum 7011 (AFSSM-7011): Emission Security Countermeasures Reviews", `http://cryptome.org/afssm-7011.htm`.

VAN BERKEL, C. H., JOSEPHS, M. B., AND NOWICK, S. M. (1999): "Scanning the Technology: Applications of Asynchronous Circuits", *Proceedings of the IEEE*, 87(2), 223–233.

VON ZUR GATHEN, J. (2003): "Claude Comiers: The first arithmetical cryptography", *Cryptologia*, 27(4), 339–349.

——— (2004): "Friederich Johann Buck: Arithmetic Puzzles in Cryptography", *Cryptologia*, 28(4), 309–324.

VTC INC. (1993): "VM7200: 2,4, 6 or 8-channel, 5-volt, thin-film head, read/write preamplifier".

WAR OFFICE (1923): "Signal Training Volume III, Pamplet No. 3: Fullerphone Mark III", HMSO, London.

WENG, Z., FALKINGHAM, C., BRIDGES, G., AND THOMSON, D. (2002): "Quantitative voltage measurement of high-frequency internal integrated circuit signals by scanning probe microscopy", *Journal of Vacuum Science and Technology A*, 20(3), 999–1003.

WESTON, D. A. (2001): *Electromagnetic Compatibility: Principles and Applications*. Marcel Dekker, New York, USA.

WRIGHT, P. (1987): *Spycatcher — The Candid Autobiography of a Senior Intelligence Officer*. William Heinemann, Australia.

YARDLEY, H. O. (1931): *The American Black Chamber*. Faber and Faber, London.

YOO, S.-K., KARAKOYUNLU, D., BIRAND, B., AND SUNAR, B. (2008): "Improving the Robustness of Ring Oscillator TRNGs", `http://ece.wpi.edu/~sunar/preprints/rings.pdf`.

YUN, J.-Y. (2001): "US Patent 6175456: Circuit for controlling the write current of a magnetic disk recording apparatus and method for optimizing the write current".

ABBREVIATIONS

3DES	*Triple-DES*
AC280	Western Digital 80 MB IDE hard drive from 1990
AEB	*ARM Evaluation Board*
AES	*Advanced Encryption Standard*
AFM	*Atomic Force Microscope*
ALU	*Arithmetic Logic Unit*
AM	*Amplitude Modulation*
AMR	*Anisotropic Magneto-Resistive*
ARC	*Authorisation Response Code*: code from EMV issuing bank to terminal to indicate whether a transaction was authorised or declined
ARM	*Advanced RISC Machines*: a microprocessor family
ARPC	*Application Response Cryptogram*: message from EMV issuing bank to terminal to authorise online transaction
ARQC	*Application Request Cryptogram*: message from EMV card to terminal to authorise transaction
ASIC	*Application Specific Integrated Circuit*
ASK	*Amplitude Shift Keying*
ATM	*Automatic Telling Machine*
ATR	*Answer To Reset*: response from ISO 7816 smartcard
CE	*Conformité Européenne*: a standards conformity mark in the European Economic Area
CGS	*Centimetre-Gram-Second*: metric system of units
CMOS	*Complementary Metal Oxide Semiconductor*
CMR	*Collossal MagnetoResistive*
DCG	*Distributed Clock Generator*

DEMA	*Differential Electromagnetic Analysis*
DES	*Data Encryption Standard*
DES	*Data Encryption Standard*
DIL	*Dual In-Line*
DPA	*Differential Power Analysis*
DSA	*Differential Spectral Analysis*
DUT	*Device Under Test*
(E)EPROM	*(Electrically) Erasable, Programmable, Read-Only Memory*
EM	*Electromagnetic*
EMC	*ElectroMagnetic Compatibility*
emf	*Electro-Motive Force*
EMI	*ElectroMagnetic Interference*
EMV	*Europay Mastercard Visa*: a standard for smartcards in banking payment systems, marketed as 'Chip and PIN' in the United Kingdom
ETB	*Emissions Testing Block*: from Lochside chip
FCC	*US Federal Communications Commission*
FIB	*Focused Ion Beam*: a post-fabrication chip editing technique
FIFO	*First In, First Out*
FIPS 140	*Federal Information Processing Standards series 140*: US government computer security standards for cryptographic modules
FIQ	*Fast Interrupt Request*: for ARM processor
FM	*Frequency Modulation*
FPGA	*Field Programmable Gate Array*
GMR	*Giant Magneto-Resistive*

GPIB	*General Purpose Interface Bus*: a parallel I/O bus for connecting and controlling laboratory instruments by computer
GSM	*Global System for Mobile Communications / Groupe Spécial Mobile*
HIJACK	Definition classified, believed to be the cross-modulation of electrical signals with sensitive information
HSM	*Hardware Security Module*
IAD	*Issuer Application Data*: data interchanged between EMV card and its issuer, whose composition is defined by the issuer and only meaningful to them
IBM	*International Business Machines Corporation*
IC	*Integrated Circuit*
IDE	*Integrated Drive Electronics*: a hard drive interfacing standard
IO	*Input/Output*
IRQ	*Interrupt Request*: for ARM processor
LED	*Light-Emitting Diode*
LFSR	*Linear Feedback Shift Register*
LH77790B	ARM7DI-based microprocessor, from 1998
MAC	*Message Authentication Code*
MEX	*MATLAB Executable*
MFM	*Magnetic Force Microscopy*: microscopy technique measuring magnetic fields to nanometre-scale spatial resolution
MFM	*Modified Frequency Modulation*: a recording technique for binary data on magnetic discs
MR	*Magneto-Resistive*
MS2601B	Anritsu spectrum analyser
NIST	*US National Institute of Standards and Technology*

NMOS	*N-type Metal Oxide Semiconductor*
NONSTOP	Definition classified, believed to be the cross-modulation of radio transmitters with sensitive information
NOP	*No Operation*
NSA	*US National Security Agency*
P7330	Tektronix differential probe for oscilloscope
PC	*Personal Computer*
PCB	*Printed Circuit Board*
PCI	*Peripheral Component Interconnect*: a PC interface bus standard
PGA	*Pin Grid Array*: a (typically ceramic) chip package using a square or rectangular array of pins
PIC	*Peripheral Interface Controller*: A popular range of microcontrollers designed by Microchip Technology
PIN	*Personal Identification Number*
PLL	*Phase-Locked Loop*
PMOS	*P-type Metal Oxide Semiconductor*
PRNG	*Pseudo Random Number Generator*
PTT	*Postes, Télégraphes et Téléphones*: French public administration of postal services and telecommunications 1921-1991
RBW	*Resolution BandWidth*
RISC	*Reduced Instruction Set Computer*
RMS	*Root Mean Square*
RNG	*Random Number Generator*
RPC	*Remote Procedure Call*
SEMA	*Simple Electro-Magnetic Analysis*

SIF	Four-wire serial interface for debugging, patented by Cambridge Consultants
SNA	*Scalar Network Analyser*
SOIC	*Small Outline Integrated Circuit*
SPA	*Simple Electromagnetic Analysis*
SPA	*Simple Power Analysis*
SPICE	*Simulation Program with IC Emphasis*: an analogue circuit simulator
SRAM	*Static Random Access Memory*
ST506	A standard for interfacing hard drives to computers, popular in the 1980s. Very similar in design to the ubiquitious Shugart SA400 floppy interface
TC	*Transaction Cryptogram*: final response from EMV card to terminal indicating that transaction was approved
TCP	*Transmission Control Protocol*
TDS2024	Portable Tektronix oscilloscope
TEAPOT	Investigation, study, and control of intentional compromising emanations (i.e., those that are hostilely induced or provoked) from telecommunications and automated information systems equipment
TEMPEST	Vulnerability of cryptographic hardware due to compromising electromagnetic emanations
TMR	*Tunnelling MagnetoResistive*
TRNG	*True Random Number Generator*
TTi	*Thurlby Thandar Instruments*
TTL	*Transistor-Transistor Logic*
UART	*Universal Asynchronous Receiver/Transmitter*: Chip for interfacing to asynchronous serial signals

UMC	*United Microsystems Corporation*: a Taiwanese semiconductor fabrication house
USB	*Universal Serial Bus/Upper Sideband*
VBW	*Video BandWidth*
VCO	*Voltage-Controlled Oscillator*
VNA	*Vector Network Analyser*
VSWR	*Voltage Standing-Wave Ratio*
VXI11	*VME eXtensions for Instrumentation 11*: an instrument control protocol over Ethernet – commands look similar to GPIB
WDI325Q	20MB MFM hard drive made in the late 1980s by IBM
XAP	Processor family designed by Cambridge Consultants Ltd
XAP1	16-bit RISC microcontroller, the first iteration of the XAP family
XOR	*Exclusive-OR*
ZIF	*Zero Insertion Force*: chip socket incorporating a clamp, so that pressure can be removed from the pins before removal/replacement

APPENDIX A
LOCHSIDE EMISSIONS TESTING BLOCK DATA SHEET

As part of the Network on Chip project, a 5mm $\times$ 5mm ASIC was fabricated in 0.18µm UMC technology. A 3mm $\times$ 0.1mm sliver of the die was available to fabricate some EMA test circuits. A number of antenna loops of differing lengths and metallisation were integrated with variable strength drivers, an on-chip clock generator including internal ring oscillators, PLLs and frequency counter. This enables a wide range of frequencies to be generated on-chip and emitted via the antennas, whose emissions can then be measured.

A.1 VERILOG DESIGN FILES

The modules of the test chip are as follows:

cam_sec_toplevel.v Root design, invoking security block and PLLs (placed in the pad ring)

cam_sec.v Top level design of security block, instantiates other modules and defines pad interface and bit-level interface to configuration registers

- **clock_gen.v** Clock generation: input clock select multiplexer
  - **clock_divider.v** Programmable 2^n clock divider
  - **counter.v** Frequency counter of divided clock
- **config_registers.v** Configuration shift register and interface to the pins
- **driver.v** Driver decode logic for variable strength drivers
  - **ant_decode.v** Decode the configuration register strength component into control wires for each drive transistor
  - **drive_logic.v** Truth table for drive of each transistor gate
- **EMA_CUSTOM_LAYOUT.v** Full custom antennas, ring oscillators and driver blocks

cam_sec_testbench.v Verilog test and verification code (only for simulation)

A.2 PIN INTERFACE

The pins of the macrocell that are accessible externally are as seen in Table A.1.

Pin name	Lochside pin	Direction	Pin function
	SECPST_in_select	I	High level selects security block, low selects pad speed test
CONFIG_CLK	SECPST_in_p0	I	Clock to configuration shift register and latching signals
CONFIG_LATCH	SECPST_in_p2	I	Enable to latch configuration from shift register and start operation
D_IN	SECPST_in_p1	I	Shift register data in
D_OUT	SECPST_out_p6	O	Shift register data out
CLK_IN	SECPST_in_p5	I	Input clock to PLL and clock generator
CLK_OUT	SECPST_out_p7	O	Output clock from clock generator if enabled
$\overline{\text{RESET}}$	SECPST_bidir_p8	I	Reset and clock shutdown
COUNT_LATCH	SECPST_in_p3	I	Enable to latch counter value into shift register
CLK_REF	SECPST_in_p4	I	Frequency counter running when high

Table A.1: Lochside ETB pin interface

The security pins are multiplexed with those of the pad speed test block. The multiplexer is supplied from the VDD of the Lochside core, whilst the security block has a separate VDD_EMA. I/O pins are driven by the USB microcontroller on the PCB, and CLK_IN, CLK_OUT and CLK_REF are connected to BNC sockets.

The device is held in reset by a low level on RESET. This clears the frequency counter, divider, and configuration registers, and initialises the PLLs. Setting all inputs to zero will keep the device in a quiescent state.

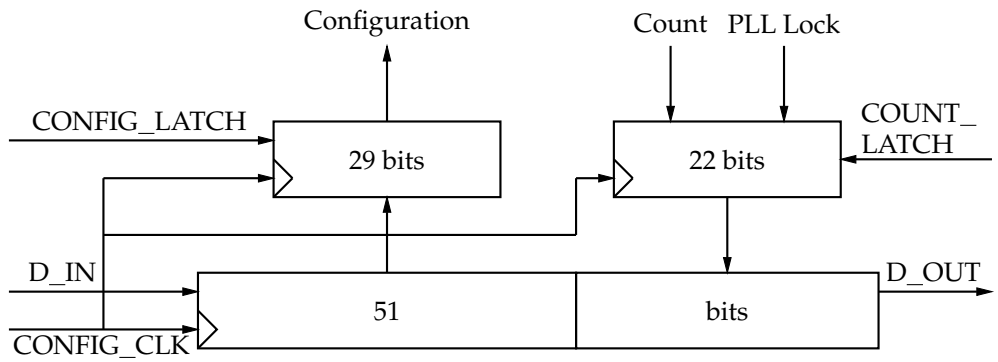


Figure A.1: Configuration shift register

A.3 CONFIGURATION INTERFACE

The device is configured via a 51-bit shift register (Figure A.1) whose data input to the least significant bit is `D_IN` and whose data output from the most significant bit is `D_OUT`, clocked off `CONFIG_CLK`. The most significant 22 bits are devoted to outputs from the device and least significant 29 bits being an input configuration word.

`COUNT_LATCH` is used to load output results from the rest of the chip into the configuration shift register, whilst `CONFIG_LATCH` is used to latch the input configuration word from the shift register into an internal register used to control all on-chip devices. These are synchronous with `CONFIG_CLK`: to load from or store data to the shift register, one or other of these signals are taken high then a rising edge on `CONFIG_CLK` provided. The shift register does not shift in this circumstance. If both `COUNT_LATCH` and `CONFIG_LATCH` are low, then the register shifts one place to the left.

The bits in the configuration control register are defined in Table A.2 on the following page (all active high unless otherwise stated).

The output register is as follows:

0—19	`count`	Frequency count
20	`pll_lock1`	Lock output of low frequency PLL
21	`pll_lock2`	Lock output of high frequency PLL

A.4 CLOCK GENERATION

The device is provided with a number of internal clock generators, an external clock input and the ability to convert any antenna into a ring oscillator. Such a clock may be divided by a power of two from 2^0 to 2^{15}, and

Bit(s)	Name	Function
0—3	clock_div	Clock divisor $2^{\texttt{clock\_div}}$
4—8	select_clock	Clock input select (see below)
9	enable_ freqcounter	Enable the frequency counter (if not enabled, the counter is held static, but not reset)
10	short_close	Close short ring oscillator loop
11	pll_bypass	Bypass second input PLL
12	long_close	Close long ring oscillator loop
13	clock_out	Emit antenna driving clock on CLOCK_OUT pin
14	enable_ask	Enable amplitude shift keying (ASK). When enabled, the antenna inputs are the logical AND of the divided clock and CLK_REF, allowing external modulation of the antenna signals.
15	enable_divider	If zero, the clock divider is forced to zero and hence is inactive. Clock signals can only propagate through the divider in this case if clock_div is zero.
16—19	select_antenna	Select output antenna
20—22	strength	Selection of antenna drive strength ($2^{\texttt{strength}} \times$ unity drive strength)
28-23	unused	Unused

Table A.2: Lochside ETB configuration bits.

output on an external pin as well as sent to an antenna.

select_clock controls which clock source is active. Note that some clock sources can be enabled even when another is being used to transmit. The clock sources are:

select_clock	Clock source
0	Direct input from CLK_IN pin
1	PLL output
2	Short ring oscillator output
3	Long ring oscillator output
4–19	End wire of antenna number (select_clock − 4)
20–31	Logical zero

To put an antenna into self oscillation mode, it is necessary to select the same antenna to drive with `select_antenna`, set `select_clock` to that value plus four and ensure that `clock_divider` is zero. Note that some antennas have outputs hardwired to ground, hence will not self-oscillate in this configuration.

The PLLs are specified as follows. The first input PLL (input from `CLK_PLL`, output to second PLL):

Design input frequency	25MHz (typ)
Permitted input frequency	14MHz — 40MHz
Design output frequency	200MHz (typ)
Duty cycle	40% (min) / 50% (typ) / 60% (max)

The second input PLL (input from output of first PLL):

Design input frequency	200MHz (typ)
Permitted input frequency	50MHz — 150MHz
Design output frequency	1000MHz (typ)
Duty cycle	40% (min) / 50% (typ) / 60% (max)

`pll_bypass` causes the output of the first PLL to be used directly and the second PLL disabled. When not selected as a clock source, both PLLs are held in reset and hence do not oscillate.

The short ring oscillator consists of 36 inverters plus a NAND gate, whilst the long ring oscillator consists of 2597 inverters (plus additional gates — see below). `long_close` and `short_close` are used to close the rings and cause them to oscillate. With these bits clear, both oscillators are forced into a known state.

A.5 ANTENNAS

The antennas are constructed as shown on the following pages, which show my schematic and actual layout designed by Andrew West. The structures are described in Table A.3 on the next page.

Metal fill is used over and between some of the antennas to meet foundry design rules. This is specified in the design file (eg see the schematic of Antenna 4), but the foundry may apply their own fill generation in addition.

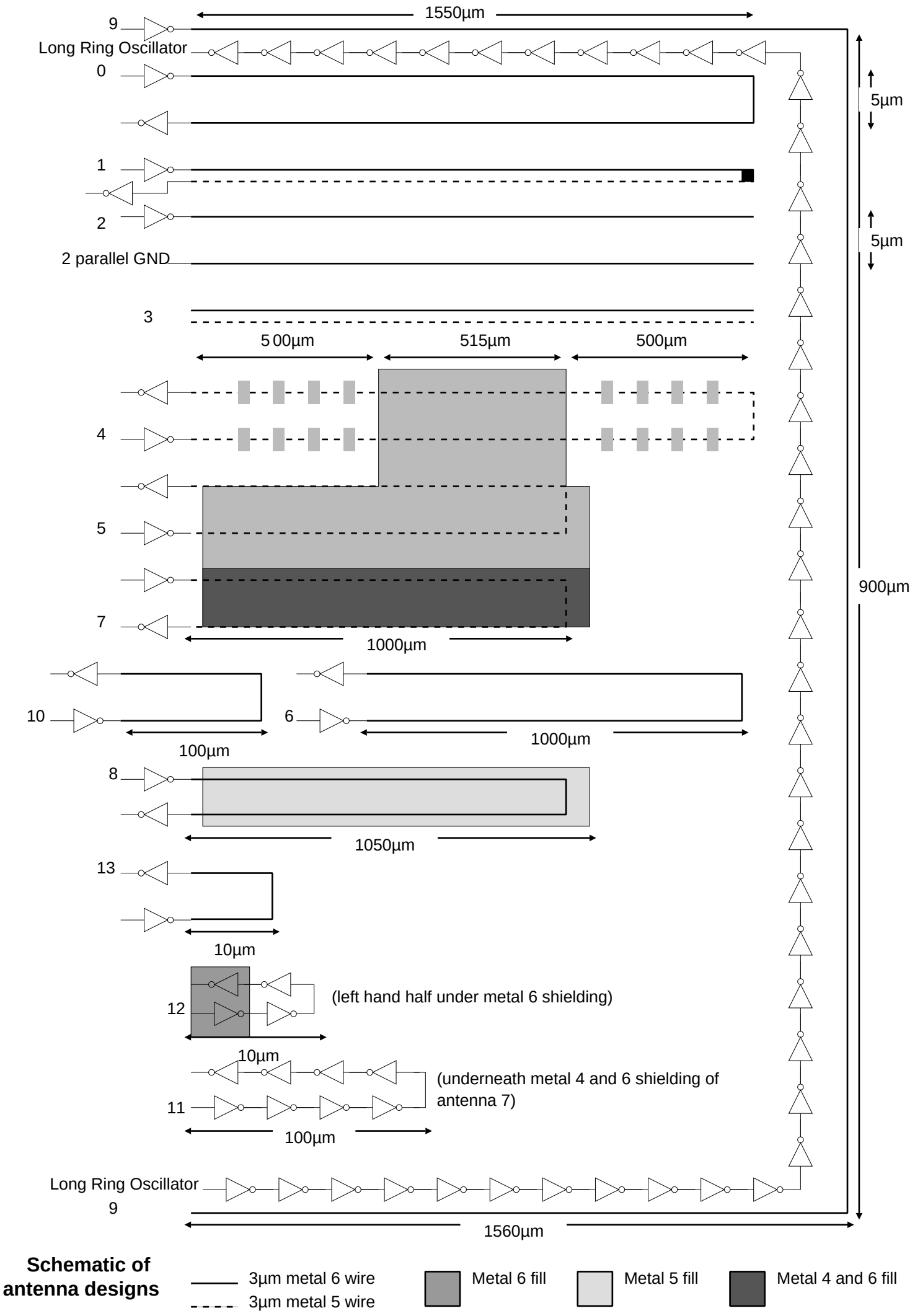
Antenna 14 selects the long ring oscillator as its output. If `long_close` is zero, the ring oscillator loop is opened and can be driven as an antenna, selected by choosing Antenna 14. If `long_close` is one, the loop is closed and driven with the logical AND of the clock output and the feedback path

Antenna	Structure
0	1.55 mm metal 6 loop separated by 5 µm
1	1.55 mm metal 6 running over return path of 1.55 mm metal 5 with connecting via at far end
2	1.55 mm parallel metal 6 wires, separation 5 µm, one driven one grounded and near end
3	1.55 mm metal 6 running over 1.55 mm metal 5 grounded at near end
4	500 µm metal 6 shield over 1.5 mm metal 5 loop, 5 µm wide.
5	1.05 mm metal 6 shield over 1.00 mm metal 5 loop, 5 µm wide
6	1 mm metal 6 loop separated by 5 µm
7	Metal 4 and 6 grounded sandwich around 1 mm metal 5 loop, 5 µm wide
8	1.05 mm metal 5 shield under 1.00 mm metal 6 loop, 5 µm wide
9	1.56 mm × 0.9 mm metal 6 loop around outside of cell
10	100 µm metal 6 loop, 5 µm wide
11	100 µm long loop of 137 inverters, between metal 4 and 6 shielding, 5 µm wide
12	10 µm long loop of 15 inverters, half under metal 6 shielding, 5 µm wide
13	10 µm metal 6 loop, 5 µm wide
14	Output from long ring oscillator

Table A.3: Lochside ETB antennas to evaluate.

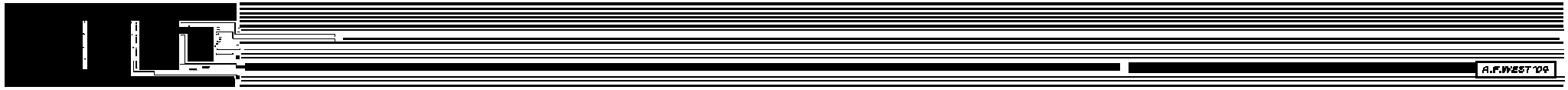
from the loop. Such behaviour may cause undefined results. If the loop is used both as an antenna and a clock source, behaviour is undefined.

Antennas are driven with a variable strength driver. This allows the drive effort applied to the antenna wire to be powers of two, between 1 and 128 times the drive strength of a single 'unity' transistor (width 2.56 µm).


1550µm
9
Long Ring Oscillator
0
5µm
1
2
5µm
2 parallel GND
3
500µm
515µm
500µm
4
5
7
900µm
1000µm
10
6
100µm
1000µm
8
1050µm
13
10µm
12
(left hand half under metal 6 shielding)
10µm
11
(underneath metal 4 and 6 shielding of antenna 7)
100µm
Long Ring Oscillator
9
1560µm
Schematic of antenna designs
3µm metal 6 wire
3µm metal 5 wire
Metal 6 fill
Metal 5 fill
Metal 4 and 6 fill

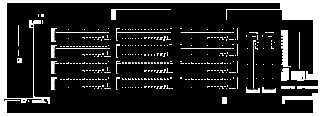
Metal 6

Metal 5



Metal 4

Metal 3



Metal 2

Metal 1

Plots of design metal layers

Dimensions 1.84mm × 0.1mm

A.6 FREQUENCY COUNTER

A frequency counter is provided. While CLK_REF is high, it counts rising edges of the antenna clock, and it halts when CLK_REF is low. CLK_REF is synchronised with the antenna clock, and hence there is a lag of two antenna clock cycles between a transition on CLK_REF and its effect on the counter. Operation when in amplitude shift keying mode is undefined. Reset of the counter is only possible with a global reset.

A.7 DISTRIBUTED CLOCK GENERATOR (DCG)

The distributed clock generator (DCG) is a completely separate functional block on the chip designed by Scott Fairbanks. It is controlled by four pins on the chip. In the host software, the three digital inputs are prepended to the 20-bit security block configuration word as seen in Table A.4, so both devices can be controlled by one 32-bit word. Further details are given in Section 4.7.2 on page 118.

Signal	Bit position in host config. word	Description
DCG_START	31	Enable the DCG oscillator
DCG_OUT_ENABLE	30	Enable the DCG output pin
DCG_DIV_CTRL	29	If 1, DCG_OUT = DCG internal frequency ÷ 16. If 0, then ÷8
DCG_SPEED_ANLG	N/A	Analogue control voltage 0 to 1.8 V

Table A.4: *DCG control inputs*