Hardware Root of Trust: QDID PUF & Attopsemi OTP

Version 1.0

Table of Contents

1 Introduction 2
1.1 Physically Unclonable Functions 2
1.2 Crypto Quantique's QDID PUF 2
1.3 Error Correction and Post processing 3
1.4 Quantum Gate Tunneling 4
1.5 One-Time Programmable 5
1.6 Attopsemi's I-fuse® OTP 5
2 Helper Data Storage 7
3 Locally Stored Helper Data 8
4 Conclusion 9
5 Revision History 10

1 Introduction

In a connected era where billions of connected devices all need to communicate securely, it is of paramount importance that each device has a unique identity and a source of high-quality randomness for cryptographic keying material. In order for services to take advantage of these connected devices and their data, each device's identity and cryptographic values (whether secret or public) needs to be shared. Many first-generation hardware security solutions have been creating identities and cryptographic keys for authentication by injecting these secret keys into each device. This process is fraught with risks: programming facilities need to be trusted or secure, increasing the cost, and keys need to be generated and stored on potentially unsafe media. There is also the risk that secret keys are leaked through human error or by intentional side channel attacks of the programmed cryptographic modules.

1.1 Physically Unclonable Functions

Physically Unclonable Functions, or PUFs, present a solution to this problem. At a high level, PUFs are physical devices made through a manufacturing process such that an individual PUF instance cannot (easily) be replicated. A PUF provides a physically generated "digital fingerprint," a sequence of high-quality random bits which can be used for cryptographic keys.[1] A PUF relies on unique physical variations which occur naturally during semiconductor manufacturing, eliminating the need to inject secret keys. Furthermore, PUFs do not require secrets to be stored in memory as the secrets are embedded within a physical function that generates the values when required. As well as mitigating the problems inherent in key injection, this has the benefit of protecting against the security risk of storing secrets in memory – as technology and reverse engineering advances, even secrets stored in protected memory are vulnerable. PUFs are increasingly being incorporated into the heart of hardware root of trust solutions

1.2 Crypto Quantique's QDID PUF

Crypto Quantique's PUF technology uses quantum effects in silicon to generate a digital fingerprint. The technology is based on the ability to measure the natural variability in gate oxide thickness produced by the manufacturing process (see Figure 1). Gate leakage currents due to quantum mechanical tunnelling in adjacent gates can be compared, and the current differences can be resolved and digitized to output either a 0 or 1. Crypto Quantique's PUF design aggregates 64x64 pairs of gates, giving an output of 4096 bits – which can then provide the source of the unique identity, we refer to as a Quantum Derived Identity (QDID). **Error! Reference source not found.**.1 shows an example of a QDID instance, where the light and dark coloured 'pixels' represent 0

[1] In the literature, the term PUF is more general and encompasses two slightly differing constructions, which we can classify as challenge/response PUFs and 'no-challenge' or 'digital fingerprint' PUFs. Challenge/response PUFs present a function that takes an input challenge and returns the same (random) output response which can be used for cryptographic purposes. A no-challenge PUF can be seen as a special type of PUF where the challenge length is 0 bits; another way of conceptualising a no-challenge PUF is as a random bit string that is physically embedded within a device. Depending on the length of the bit string, this can be divided up and used for cryptographic secrets.

and 1 respectively. Due to inherent instability in the design process, the leakage currents are unpredictable, and inherently random with a high entropy.

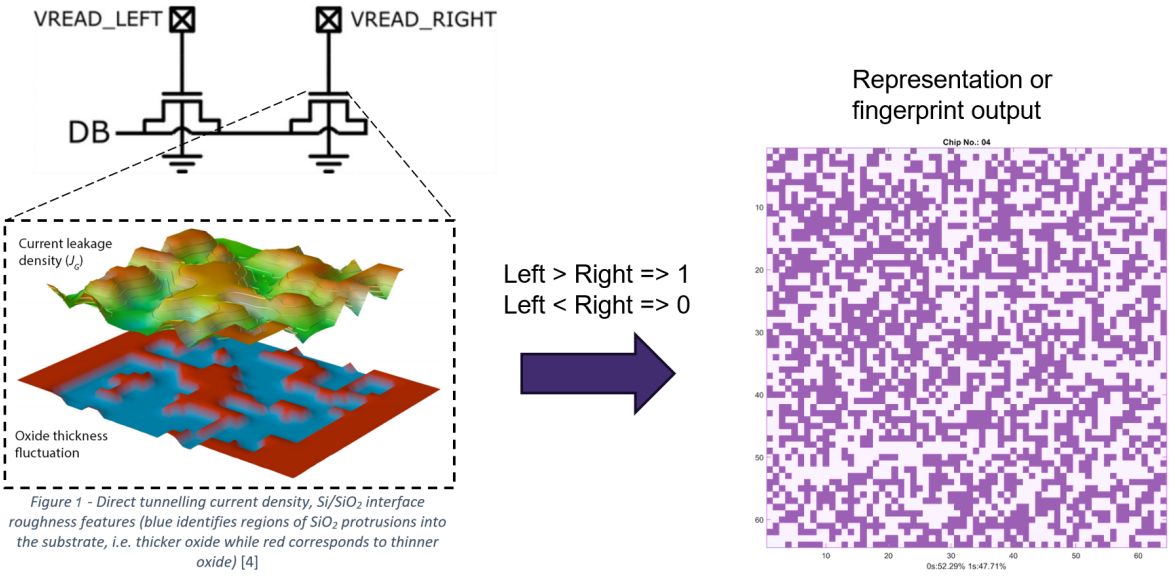

Figure 1 - Direct tunnelling current density, Si/SiO_2 interface roughness features (blue identifies regions of SiO_2 protrusions into the substrate, i.e. thicker oxide while red corresponds to thinner oxide) [4]

Figure 1-1: A unique, unclonable and tamper-evident QDID

Each time a readout is taken, the seed will be fractionally different from a reference read. Post-processing ensures that the errors are corrected for and the true value is used for consistent cryptographic keys. This has been introduced in the section below.

1.3 Error Correction and Post processing

Outputs of the QDID PUF are subject to errors which need to be corrected to output the same key consistently. One approach to eliminate these errors is to use an error correcting code. In abstract terms, error correction takes an input string, corrects any errors and outputs a corrected string (which will be used to derive a key). It will output the same corrected string for every input string that is 'close enough' to the reference string. 'Close enough' means with an error that is not too big; in practice, anything up to 10% bit flips is easily manageable – certainly, the low error rates demonstrated for QDID cause no problems for the process of error correction.

Error correcting codes are typically designed to correct transmission errors over unreliable channels. Cyclic error correcting codes, such as BCH (Bose-Chaudhury-Hocquenghem) require some values, called "Helper Data" to be stored in memory. During a key reproduction, the error correction codes retrieve the helper data and a PUF array read to produce an error corrected secret key or seed.

Error correction has the property that the cryptographic keys generated preserve the entropy of the QDID PUF readout. In other words, the output of post processing (cryptographic key) has good entropy if its input (fingerprint readout from QDID) has good entropy. Some entropy is lost through the error correction, proportional to the level of error. In other words, there is a cost to pay in entropy to correct unreliable inputs. In practice what this means is that we need to use an *n*-bit input string to make a reliable

key with *k*-bits of entropy (with *n > k).* For our error rates, we can choose the parameters for post-processing to ensure that generating keys will fail with a probability of 10^{-9}.

1.4 Quantum Gate Tunnelling

As predicted by the International Technology Roadmap for Semiconductors, aggressive scaling of Complementary Metal Oxide Semiconductor (CMOS) devices has resulted in gate oxide thicknesses beginning to reach the level of just tens of atoms. For thicknesses approaching a few nanometres, gate leakage currents are non-negligible because of the quantum tunnelling of carriers through the gate dielectric. The physical mechanisms which induce gate leakage current depend on the oxide thickness, potential difference across the oxide, and electrical and thermal stress conditions.

Figure **Error! Reference source not found.** illustrates the concept of energy band diagrams for different materials which align to the free electron level. The free electron level is defined as the energy level where an electron is no longer bonded to the lattice. In the figure, E_C is the conduction band edge, E_V is the valence band edge, E_i is the intrinsic Fermi level and E_F is the Fermi level which is close to the valence band edge for doped p-type silicon and close to the conduction band for doped n-type polysilicon representative of a NMOS transistor. Part (d) of the figure shows the main leakage mechanisms in ultra-thin silicon dioxide. These are Fowler-Nordheim tunnelling (FNT), Quantum Mechanical Direct Tunnelling (QMDT) and Trap Assisted tunnelling (TAT). The net gate leakage current is the summation of all three and the dominant contributor is dependent on the applied gate-to-source voltage.

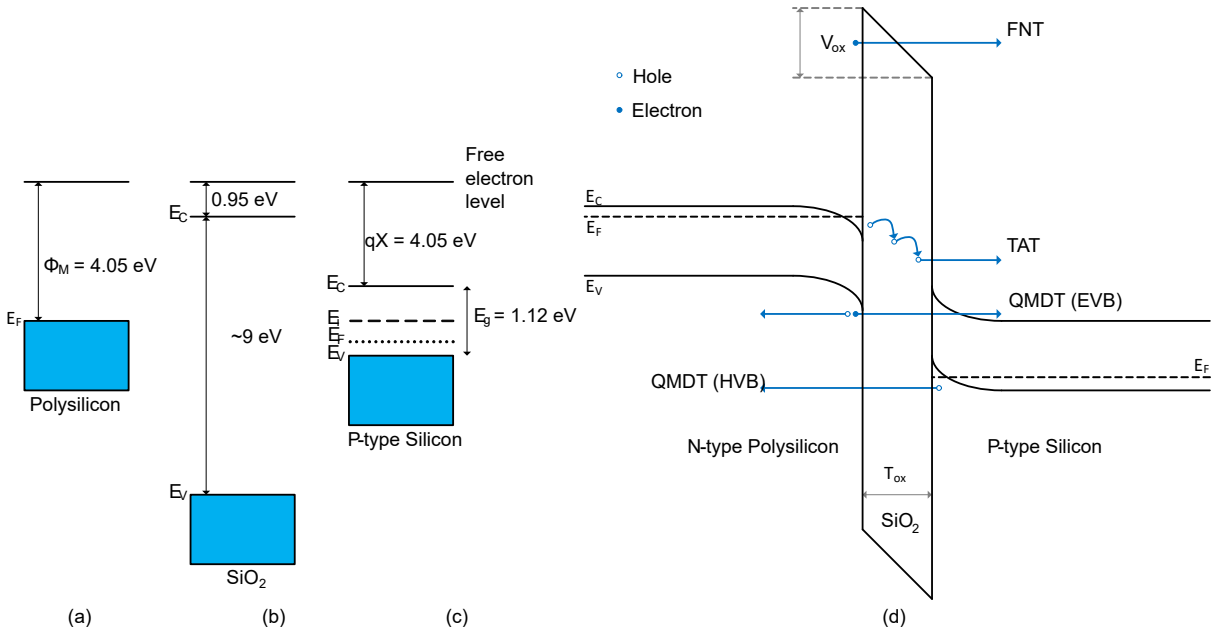

Figure 1: Energy band diagrams for the three components when separated: (a) the polysilicon gate; (b) the silicon dioxide gate; (c) the silicon doped wafer; (d) the three energy components when brought together in the accumulation regime showing FNT, QMDT and TAT physical mechanisms.

The physical mechanisms described by the energy band theory leads to the identification of 5 gate tunnelling components within a MOS transistor associated with the source, drain and substrate, as shown in Figure **Error! Reference source not found.**.

In Figure 3, I_{gst} and I_{gdt} denote the tunnelling currents between the gate-to-source and gate-to-drain regions due to the overlap region as a result of lateral diffusion of the source and drain implants; I_{gcs} and I_{gcd} are the gate-channel-source and gate-channel-drain tunnelling currents, while I_{gbt} denotes the tunnelling current between the gate and substrate.

Studies indicate that the I_{gcs} and I_{gcd} are the dominant mechanisms because of the ultra-thin gate oxide and high electric field strength. The I_{gst} and I_{gdt} currents are secondary as the overlap area is much smaller owing to the lateral diffusion of the N^+ implants and the width of the device with respect to the total area under the channel.

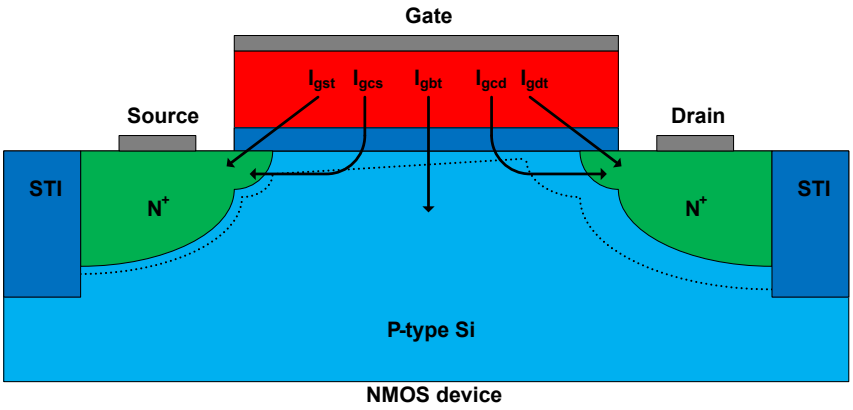

Figure 3: Device cross-section showing the gate-to-source, gate-to-substrate and gate-to-drain leakage paths present for ultra-thin gate oxides.

1.5 One-Time Programmable

One-Time Programmable, or OTP, a device that can only be programmed once to store data permanently but ideally read infinite times. OTP is used in many applications, for example, to customize a chip after fabrication, to store chip ID, firmware, security key, or configuration data, work around defect/contamination, trim device variations, or enable/disable certain functions. OTP is one of the four foundational IPs, along with the I/O library, standard cell library, and SRAM compiler.

1.6 Attopsemi's I-fuse® OTP

Attopsemi's I-fuse® OTP is a revolutionary non-breaking fuse technology that can be reliably programmed by heat assisted electromigration below a break point. Its innovative programming mechanism supports a wide voltage range, allowing for in-field programming, smaller silicon area without charge pumps and rapid programming speeds. For instance, Attopsemi's 8K-bit OTP design supports a programming voltage of 2.1V and allows single-bit programming in just 10 µs, with a compact IP size approximately 0.0279mm^2. Additionally, the I-fuse® OTP is based on a pure logic

process which ensures its compatibility across various foundry processes and applications.

2 Helper Data Storage

As discussed in section 1.3, prior to seed reproduction, the QDID PUF requires the Helper Data to be pre-loaded so that the raw data from the QDID array can be corrected for any errors. The Helper Data is required to be stored in the system for retrieval during any key (seed) reproduction. Typically, it is stored in non-volatile memory which is typically readily available in microcontroller or System-On-Chip (SoC) systems (see Figure 4).

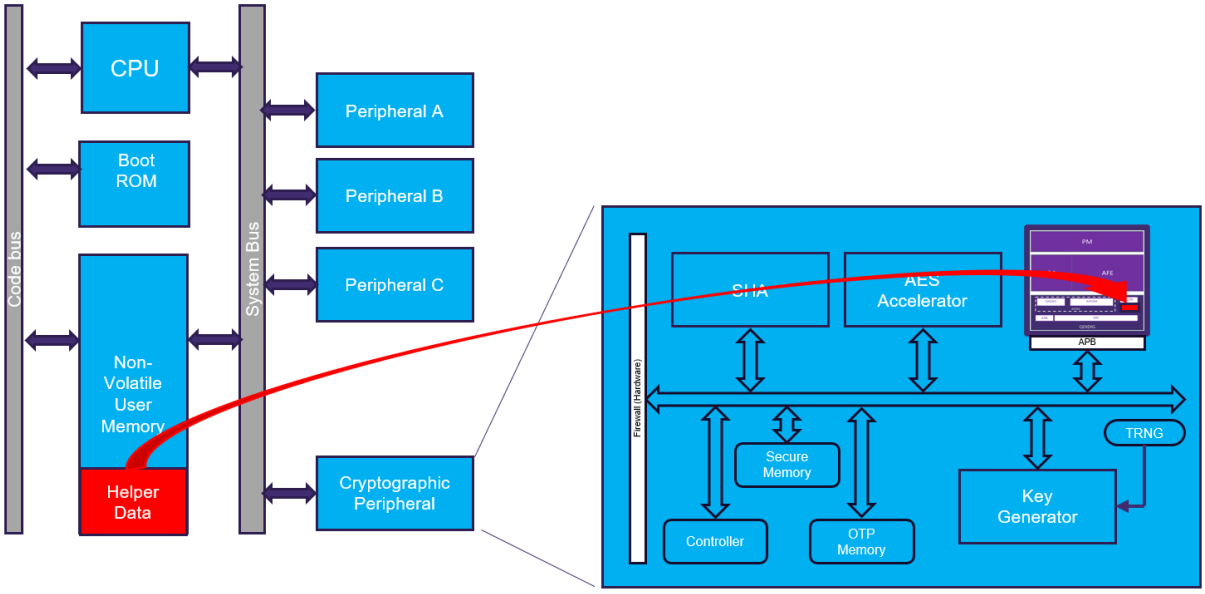

Figure 4: Typical SoC/MCU Block Diagram

*MCU - Microcontroller

There are several disadvantages to the storage of the Helper Data in non-volatile memory:

1. Helper Data, although small, takes up non-volatile memory resources that may be required by the system for other functions.
2. A writing sequence to the non-volatile memory is required during the QDID PUF initialisation within the system. This requires that the system includes non-volatile memory write functions which are often complex due to the requirement to handle power supply interruptions.
3. Any corruption of the Helper Data by erroneous writes to non-volatile memory, by other system functions that have access, will prevent re-production of QDID seeds/keys, resulting in system failure.

An alternative solution is suggested in the next section.

3 Locally Stored Helper Data

A practical consideration for the error correction implementation is in the storage of the Helper Data during the initial key generation phase. The SoC/MCUs generally access the PUF Helper Data and seed through the interface bus (see Advanced Peripheral Bus (APB) in Figure 4). These could be a dedicated secure bus to ensure protection against side-channel leakage through a generic bus or shared interface circuits. To reduce the attack surface, the PUF solution would benefit from an integrated secure One-Time-Programmable (OTP) which would allow the storage (during seed generation) and retrieval (during seed reproduction) of the helper data controlled internally within the root of trust as highlighted in Figure 5.

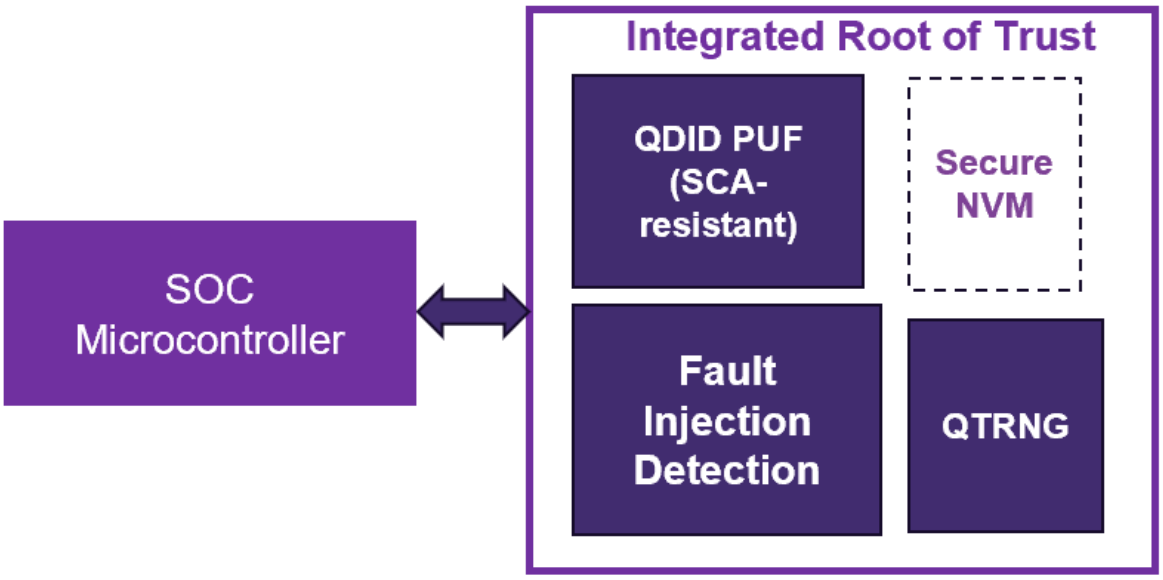

Figure 5: Architecture for integrated Helper Data storage

There are many advantages to the architecture shown in Figure 5. Crypto Quantique have teamed up with Attopsemi® and developed an integrated solution that allows the storage of Helper Data locally to the QDID PUF. Attpsemi's I-fuse® technology is foundry independent which ensures its compatibility with QDID PUF. The technology also includes a flexible programming interface and compatible power supplies which greatly simplifies the integration process. Importantly, storing the Helper Data in OTP technology local to the QDID PUF provides the following advantages:

1. Helper Data does not take up SoC/MCU resources.
2. Helper Data is written to the OTP during the die manufacturing test phase. No software, programming process or initialisation needs to be executed by the SoC/MCU user.
3. SoC/MCU user does not have access to the OTP that is storing the Helper Data. No corruption or helper data manipulation is possible during normal operation of the SoC/MCU.
4. Helper Data registers do not need to be made visible to the SoC/MCU user.

5. The integration of the Crypto Quantique QDID PUF and Attopsemi® I-fuse OTP offers the most compact root of trust solution for secret key generation within SOCs.

Figure 6 shows the architecture of the integrated QDID PUF and OTP technology.

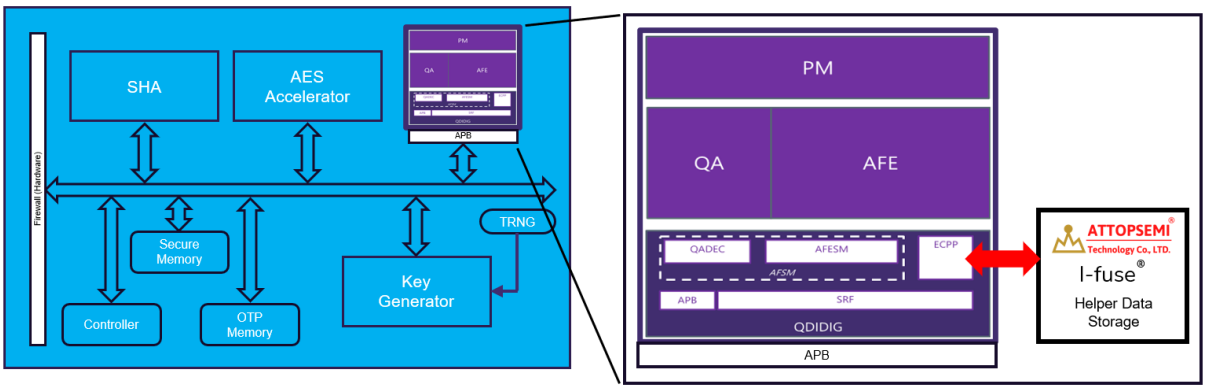

Figure 6: Integrated QDID PUF and Attopsemi® technology

4 Conclusion

Integration of Attopsemi® I-fuse® technology with the QDID PUF significantly simplifies the end user experience when using PUF technology in a root-of-trust for a SoC/MCU. The initialisation process of the QDID PUF is removed for the end user as this is implemented during manufacturing. The solution eliminates the need for the end user to manage Helper Data, and it does not require any description or special considerations in the documentation. Seed (keys) are simply read from the QDID PUF when required. There is no latency increase and side channel protection is improved due to the use of I-fuse®.

Crypto Quantique and Attopsemi® technology integration has significantly improved the overall user experience of a quantum-based PUF with additional improvements in security and side channel attack immunity.

5 Revision History

Rev.	Date	Owner	Description
Draft 5	22.09.2024	CDJ/RD/BJ	

CRYPTO QUANTIQUE

United Kingdom

Unit 304-5,
164-180 Union Street,
London
SE1 0LH

General contact email:
info@cryptoquantique.com