Security analysis of Quantum-Readout PUFs in the case of challenge-estimation attacks

Boris Škorić

Abstract

Quantum Readout PUFs (QR-PUFs) have been proposed as a technique for remote authentication of objects. The security is based on basic quantum information theoretic principles and the assumption that the adversary cannot losslessly implement arbitrary unitary transformations on a K-dimensional state space, with K 'large'. We consider all possible attacks in which the adversary bases his response on challenge state estimation by measurements. We first analyze the security of QR-PUF schemes in the case where each challenge consists of precisely n identical quanta. We use a result by Bruß and Macchiavello to derive an upper bound on the adversary's success probability as a function of K and n. Then we generalize to challenges that contain a probabilistic number of quanta, and in particular a Poisson distribution.

1 Introduction

1.1 Physical Unclonable Functions

Authentication is usually based on either "something that you know" or "something that you possess". In the second case it is desirable to work with tokens that are difficult to clone, even for the manufacturer of the token. With the advent of Physical Unclonable Functions (PUFs), physical systems have been developed which satisfy strong uniqueness and unclonability properties, e.g. phenomena such as laser speckle based on multiple scattering. A PUF is a complex piece of material whose structure is difficult to reproduce accurately because its manufacture contains uncontrollable steps [1, 2, 3, 4, 5, 6, 7, 8, 9]. A stimulus can be applied to the PUF ('challenge'), leading to a 'response' that depends in a complex way on the challenge and the precise details of the PUF's structure. The combination of a challenge and the corresponding response is called a Challenge-Response Pair (CRP).
An example of a physical system satisfying the above requirements is the so-called Optical PUF: a three-dimensional diffusive structure containing randomly positioned optical scatterers. When an Optical PUF is illuminated by a laser, the transmitted and reflected light has a random-looking pattern of dark and bright spots known as speckle. The characteristics of the laser beam (e.g. wavelength, angle, focus) constitute the challenge; the speckle pattern is the response. The response depends strongly on the challenge and on the exact positions of the scatterers. Optical PUFs support a large number of independent CRPs. [10, 11, 12]

1.2 Quantum readout of PUFs

A PUF-based authentication or anti-counterfeiting system typically has two phases: enrollment and verification. In the enrollment phase the Verifier applies a limited number of random challenges to a PUF and records the CRPs in a database. Later, in the verification phase, the Verifier has to decide whether a PUF is authentic. He looks up the CRPs listed for that given PUF, and by challenging the PUF anew verifies if it produces the listed responses. The procedure sketched above is extremely reliable when the Verifier has full physical control over the PUF. There are many cases, however, where the PUF owner is unwilling or unable to hand over his PUF. In

such situations the Verifier must do verification without having full control. This is referred to as "hands-off" verification. Achieving a high level of security is far more difficult in this setting, since there is a serious danger of emulation ('spoofing').
In practical situations, the number of supported independent CRPs is 'small' in the sense that anyone holding the PUF can, in a feasible amount of time, extract enough information from the PUF to be able to compute (or look up) the response to any future PUF challenge without having to use the PUF any more. Thus we must assume that a PUF can be *emulated* once the adversary has had a chance to examine it. This also holds for Optical PUFs, though the emulation may require quite a large database of CRPs. In general, the stricter the robustness requirements (i.e. reproducibility of responses), the smaller the challenge space and hence the more serious the danger of emulation.
The usual way to retain control in the "hands-off" setting is to have a trusted measurement device in the field or extra sensors for detecting specific kinds of spoofing. This approach has a drawback: The extra anti-spoofing hardware adds cost, while it is difficult to ascertain how secure the system actually is. For instance, remote trusted devices need to be tamper-proofed, but hardware attacks improve over time. Similarly, new techniques are continuously developed to spoof sensors. Thus, it is an arms race situation.
An elegant way out of this expensive arms race was proposed in [13]: Quantum Readout (QR) of PUFs. It makes spoofing fundamentally difficult by making use of basic quantum information theoretic principles. The main idea is to have PUF challenges that are quantum states, so that the adversary cannot extract all information from them; if he does not know the challenge, he does not know what to emulate. This approach is fundamentally secure as long as the adversary does not have the means to efficiently[1] apply arbitrary unitary transformations to the quantum state.
More specifically, the scheme works as follows. The PUF interacts with the challenge state via unitary evolution and produces a response that is also a quantum state. The Verifier, who knows from the enrollment phase what the response state is supposed to be, is able to verify if the response is correct. All this can be done without a trusted remote device, because of the inherent tamper-resistant properties of single quanta. The No Cloning Theorem [16, 17] ensures that an unknown single quantum state cannot be copied onto another particle. One of the implications is that the state of an unknown quantum challenge cannot be fully determined. By repeatedly sending random challenges, the verifier ensures that the probability of successful spoofing is brought down exponentially. Nice properties of the QR-PUF technique are that the challenge space does not have to be large, and that the scheme is still secure if the list of responses is publicly known.
Quantum Readout of PUFs was first experimentally realized by Goorden et al. [15] in an Optical PUF system. The challenge was implemented as a weak coherent light pulse with average photon number n_{av} and a randomly chosen wavefront that has K degrees of freedom, with $K \gg n_{\mathrm{av}}$. The scattering in the PUF scrambles the wavefront. The response is the scrambled light pulse. Verification was performed using a spatial light modulator and a photon counter. The security is based on the fact that performing measurements on n_{av} photons (give or take a few) reveals too little information to characterize the K-mode challenge state.

1.3 Security of Quantum-Readout PUFs: previous work

The existing security analyses of QR-PUFs assume that the adversary does not have a way to perform arbitrary unitary operations. The analyses are restricted to so-called *challenge estimation* attacks, in which the adversary first does a measurement on the challenge, from the outcome calculates an estimate of the challenge and finally produces a response quantum state consistent with the estimated challenge. We will also work in this context.
In [14] it was shown for the case of single-quantum challenges that the per-round false accept probability cannot exceed $2/(K+1)$, where K is the dimension of the Hilbert space. Hence, QR-PUFs can be secure against challenge estimation even if the dimension of the Hilbert space is low, e.g. $K = 2$.

[1] By 'efficient' we mean without particle losses, fast, and at a reasonable cost. At the moment, and in the foreseeable future, there is no lossless way to apply arbitrary unitary operations in the optical PUF system of [15].

Ref. [18] analyzed the case of Optical QR-PUFs [15] with K modes and average photon number n_{av}. The adversary has (on average) n_{av} quanta to examine, which gives him more information than in the single-quantum case. A specific type of measurement was considered, known as *quadrature*, which is the most informative kind of measurement known for electromagnetic fields. It was shown that a challenge estimation attack in this context cannot achieve a per-quantum false accept probability better than approximately $n_{\text{av}}/(K + n_{\text{av}})$. Thus, security is achievable as long as n_{av} is not large compared to K.
In the case of multiple-quantum challenges, existing security analyses of QR-PUFs do not provide a result for *generic* challenge estimation attacks, i.e. attacks without any assumption about the measurements performed by the adversary.

1.4 Contributions

We analyze the security of QR-PUFs against generic challenge-estimation attacks, where we allow the adversary to do arbitrary measurements such as Positive Operator-Valued Measures (POVMs). We consider two cases: (i) The challenge comprises exactly n quanta. (ii) The number of quanta is not fixed, but follows a probability distribution; in particular we focus on coherent states.

- For fixed n, we derive an upper bound $\frac{n+1}{n+K}$ on the adversary's per-quantum accept probability. This bound follows directly from a recent result by Bruß and Macchiavello [19]. Bounds on the per-round accept probability and on the overall false accept probability follow straightforwardly.
- For a distribution of the number of quanta with average n_{av}, we show that the per-quantum accept probability is upper bounded by $\frac{n_{\text{av}}+1}{n_{\text{av}}+K}$.
- We derive a closed-form upper bound on the per-quantum accept probability in the case where the challenge is a coherent state. The result does not differ much from the generic bound $\frac{n_{\text{av}}+1}{n_{\text{av}}+K}$.

2 Preliminaries

2.1 Notation

Quantum states are represented as vectors in a Hilbert space. We adopt the usual Dirac 'bra' and 'ket' notation; the ket vector $|\psi\rangle$ stands for a quantum state labelled by some description ψ which summarizes all the knowable information about the state. The Hermitian conjugate is denoted as the bra vector $\langle\psi|$. The notation for the inner product between two states is $\langle\psi_1|\psi_2\rangle$. We will consider only normalized states, i.e. satisfying $\langle\psi|\psi\rangle = 1$. We will consider a K-dimensional Hilbert space. The properties of the PUF are summarized as a unitary $K \times K$ transition matrix R. The PUF response to a challenge $|\psi\rangle$ is $R|\psi\rangle$.

2.2 Attacker model

We consider the following attacker model. The verifier prepares a challenge consisting of exactly n quanta (with $n < K$) that are all in the same pure state $|\psi\rangle$. The state $|\psi\rangle$ is chosen uniformly at random. He sends the challenge to the PUF holder. There the challenge interacts with the PUF, resulting in a response state. The challenge state can be written as $|\Psi\rangle = \otimes_{\alpha=1}^{n}|\psi\rangle_\alpha$, and the expected response state is $|\Omega\rangle = \otimes_{\alpha=1}^{n}|\omega\rangle_\alpha$ with $|\omega\rangle = R|\psi\rangle$. The response state is returned to the verifier. The protocol has N rounds.
For each quantum in $|\Omega\rangle$ independently the verifier checks the validity of the response. He does this by projecting each response quantum onto $|\omega\rangle$ (with measurement outcome 1 in the case of a match and 0 otherwise). We assume that he has the technological means to measure the projection operator $|\omega\rangle\langle\omega|$ for arbitrary $|\omega\rangle$. Ideally, the correct response yields n matches. However, imperfections in the equipment may cause some noise. (Noise can occur at any stage: challenge

preparation, state transport, interaction with the PUF, and measurement of $|\omega\rangle\langle\omega|$.) In order to accommodate for such noise, the verifier tolerates a fraction $\varepsilon_{\text{noise}}$ of all projection outcomes (i.e. over multiple rounds in the protocol) to be zero.
We investigate the following attack. The adversary fully knows R but does not possess the PUF. Furthermore, he does not possess a quantum computer or, equivalently, a device that can perform arbitrary unitary operations in a lossless way. The adversary performs a generic measurement on $|\Psi\rangle$, described by a POVM, in order to estimate $|\psi\rangle$ as accurately as he can. We denote his estimate as $|\hat{\psi}\rangle$. He computes $|\hat{\omega}\rangle = R|\hat{\psi}\rangle$, prepares this state n times and sends $|\hat{\Omega}\rangle \stackrel{\text{def}}{=} \otimes_{\alpha=1}^{n}|\hat{\omega}\rangle$ back to the verifier.
We say that the attack has succeeded if, in the whole set of challenge-response rounds, the success rate exceeds $1 - \varepsilon_{\text{noise}}$.

2.3 State estimation

In our context, the most natural figure of merit to express the 'goodness' of the adversary's estimate $|\hat{\psi}\rangle$ is the *fidelity*, which is defined as

$$F \stackrel{\text{def}}{=} \mathbb{E}_{\psi}|\langle\hat{\psi}|\psi\rangle|^2, \tag{1}$$

where $\mathbb{E}_{\psi}$ stands for the expectation over the randomly chosen state $|\psi\rangle$. Note that $|\hat{\psi}\rangle$ is a function of $|\psi\rangle$, the choice of measurement, and the (probabilistic) outcome of the measurement.

Lemma 1 (from [19]) *The maximum achievable fidelity for state estimation from n identical copies of a K-dimensional quantum system is $F_{\max} = \frac{n+1}{n+K}$.*

3 Security analysis

We determine the attacker's success probability in the model specified in Section 2.2.

3.1 The False Accept probability per quantum

For each of the attacker's response quanta independently there is a probability $P_{\text{accept}|\psi}$ that the state $|\hat{\omega}\rangle$ will be projected to $|\omega\rangle$,

$$P_{\text{accept}|\psi} = |\langle\omega|\hat{\omega}\rangle|^2 = |\langle\psi|R^{\dagger}R|\hat{\psi}\rangle|^2 = |\langle\psi|\hat{\psi}\rangle|^2. \tag{2}$$

Here we have used the fact that R is unitary, i.e. $R^{\dagger}R = \mathbf{1}$. The probability of single-quantum correct projection is

$$P_{\text{accept}} = \mathbb{E}_{\psi}P_{\text{accept}|\psi} = F. \tag{3}$$

Theorem 1 *In the QR-PUF protocol and attack as specified in Section 2.2, the adversary's probability P_{accept} that a response quantum is accepted by the verifier satisfies*

$$P_{\text{accept}} \leq \frac{n+1}{n+K}.$$

Proof: Follows directly from (3) and Lemma 1. □

3.2 False Accept probability of the protocol

The expected number of passing attacker quanta in one round is nP_{accept}, and the expected total number in the protocol is NnP_{accept}. Hence, if P_{accept} exceeds $1 - \varepsilon_{\text{noise}}$ then the protocol will typically accept the attacker.

The precise False Accept probability is computed as follows. Since the projections of the quanta are independent events, the number of accepted quanta is binomial-distributed. We denote the number of accepted quanta as A, with distribution

$$p_{Nn,k} \stackrel{\text{def}}{=} \Pr[A = k] = \binom{nN}{k} P_{\text{accept}}^k (1 - P_{\text{accept}})^{nN-k}. \tag{4}$$

The probability P_{FA} that the adversary successfully completes the authentication protocol is given by

$$P_{\text{FA}} = \sum_{k=\lceil Nn(1-\varepsilon_{\text{noise}})\rceil}^{Nn} p_{Nn,k}. \tag{5}$$

3.3 Probabilistic number of quanta

We investigate what happens if the number of quanta in the challenge is not fixed. For instance, the number of photons in the coherent challenges of [15] is Poisson-distributed. Another example is a cheap single-photon source which occasionally creates multiple photons instead of a single one. The protocol needs only one minor modification. Typically the verifier does not know n, but only its distribution. The verification threshold can be set to $Nn_{\text{av}}(1-\varepsilon)$, where ε must take into account not only noise but also the variance of n.

Theorem 2 *Let n, the number of quanta in the challenge, have a probability distribution with mean n_{av}. The single-quantum accept probability for the attacker is then bounded as*

$$P_{\text{accept}} \leq \frac{n_{\text{av}} + 1}{n_{\text{av}} + K}. \tag{6}$$

Proof: We have to average the acceptance probability over the number of quanta n. This gives $P_{\text{accept}} \leq \mathbb{E}_n \frac{n+1}{n+K} = 1 - (K-1)\mathbb{E}_n \frac{1}{n+K}$. We use Jensen's inequality for convex functions and get $\mathbb{E}_n \frac{1}{n+K} \geq \frac{1}{\mathbb{E}_n n+K} = \frac{1}{n_{\text{av}}+K}$. □

Theorem 2 holds for arbitrary distributions. The result below is specific for the Poisson distribution.

Theorem 3 *Let the number of quanta in the challenge be Poisson-distributed with mean n_{av}. The single-quantum accept probability for the attacker is then bounded as*

$$\begin{aligned} P_{\text{accept}} &\leq 1 - (K-1)e^{-n_{\text{av}}}(-n_{\text{av}})^{-K}[\Gamma(K) - \Gamma(K, -n_{\text{av}})] && (7)\\ &= 1 - \frac{(K-1)(K-1)!}{(-n_{\text{av}})^K}\left[e^{-n_{\text{av}}} - \sum_{b=0}^{K-1}\frac{(-n_{\text{av}})^b}{b!}\right], && (8) \end{aligned}$$

where $\Gamma(\cdot,\cdot)$ is the incomplete Gamma function. If $n_{\text{av}} < K$ then the bound can be written as

$$\begin{aligned} P_{\text{accept}} &\leq \frac{1}{K} + \sum_{a=1}^{\infty}(-1)^{a+1}(\frac{n_{\text{av}}}{K})^a(1-\frac{1}{K})\frac{K^a}{(K+1)_a} && (9)\\ &= \frac{n_{\text{av}}+1}{n_{\text{av}}+K} - \sum_{a=1}^{\infty}(\frac{n_{\text{av}}}{K})^a(-1)^{a+1}(1-\frac{1}{K})[1 - \frac{K^a}{(K+1)_a}] && (10)\\ &= \frac{n_{\text{av}}+1}{n_{\text{av}}+K} - \frac{1}{K}\left[\frac{n_{\text{av}}}{K} - \frac{3n_{\text{av}}^2 + 2n_{\text{av}}}{K^2} + \mathcal{O}(\frac{n_{\text{av}}^2 + n_{\text{av}}}{K^3})\right], && (11) \end{aligned}$$

where $(K+1)_a = (K+a)!/K!$ is the Pochhammer symbol.

Proof: We have $P_{\rm accept}^{\rm max} = \mathbb{E}_n[1 - \frac{K-1}{n+K}] = 1 - \sum_{n=0}^{\infty} p(n)\frac{K-1}{n+K}$, where $p(n) = e^{-n_{\rm av}} n_{\rm av}^n/n!$ is the Poisson distribution. We write

$$P_{\rm accept}^{\rm max} = 1 - (K-1)e^{-n_{\rm av}} n_{\rm av}^{-K} A(n_{\rm av}, K), \quad \text{with} \quad A(n_{\rm av}, K) = \sum_{n=0}^{\infty} \frac{n_{\rm av}^{n+K}}{(n+K)n!}. \tag{12}$$

We note that $A(0, K) = 0$ and that

$$\frac{\partial}{\partial\lambda} A(\lambda, K) = \sum_{n=0}^{\infty} \frac{\lambda^{n+K-1}}{n!} = \lambda^{K-1} e^{\lambda} \sum_{n=0}^{\infty} p(n) = \lambda^{K-1} e^{\lambda}. \tag{13}$$

This allows us to write

$$A(n_{\rm av}, K) = \int_0^{n_{\rm av}} {\rm d}\lambda\, \lambda^{K-1} e^{\lambda} = \int_{-n_{\rm av}}^{0} {\rm d}x\, (-x)^{K-1} e^{-x} = (-1)^{K-1}[\Gamma(K, -n_{\rm av}) - \Gamma(K)]. \tag{14}$$

Substitution into (12) yields (7). The integration can also be evaluated via integration by parts $K-1$ times,

$$\int_0^{n_{\rm av}} {\rm d}\lambda\, \lambda^{K-1} e^{\lambda} = (-1)^{K-1}(K-1)! \left[e^{\lambda} \sum_{b=0}^{K-1} \frac{(-\lambda)^b}{b!} \right]_{\lambda=0}^{n_{\rm av}}. \tag{15}$$

For $\lambda = 0$, the expression between square brackets evaluates to 1. Substitution of (15) into (12) yields (8). Eq. (9) directly follows from (8) by organizing the expression in powers of $n_{\rm av}$. Eq. (10) follows by subtracting the Taylor expansion of $\frac{n_{\rm av}+1}{n_{\rm av}+K}$ in the small parameter $n_{\rm av}/K$. Keeping only the first two terms of the summation in (10) yields (11). □

From (11) we see that the difference between the bounds in Theorem 2 and 3 is minute, unless K is small and $n_{\rm av}$ is not small compared to K.
Fig. 1 shows the bound (7) as a function of $n_{\rm av}$ and K. Fig. 2 shows the difference between Theorem 2 and 3 for $K = 20$. It is visible but small, and becomes smaller with increasing K. ($K = 20$ is on the low side, given that we need to satisfy the security assumption that the adversary cannot losslessly realize arbitrary unitary operations. See Section 2.2.)

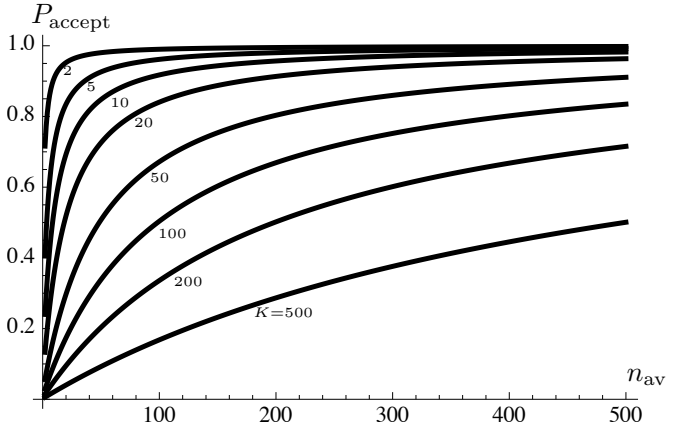

Figure 1: *According to Theorem 3: upper bound on $P_{\rm accept}$ as a function of $n_{\rm av}$, for different K, in case of the Poisson distribution.*

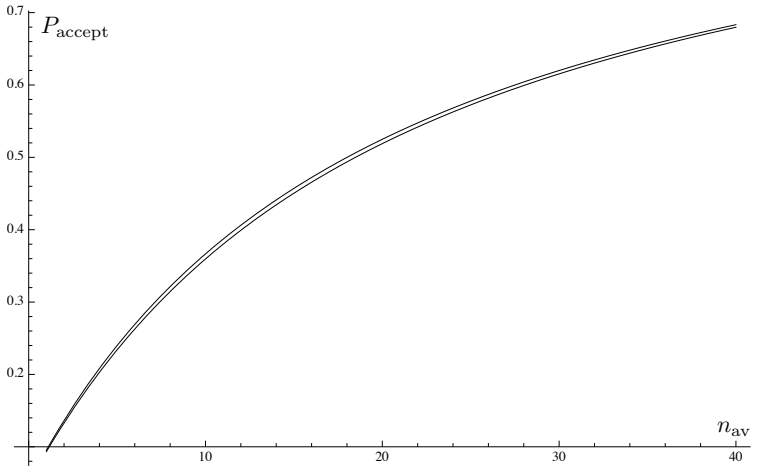

Figure 2: *The bound on the attacker's single-quantum acceptance probability, according to Theorem 2 (upper curve) and Theorem 3 (lower curve), for* $K = 20$.

4 Discussion

We have made use of a fundamental result by Bruß and Macchiavello [19] to bound the success probability of a challenge estimation attack on QR-PUF authentication. Their bound $F \leq \frac{n+1}{n+K}$ on the fidelity immediately leads to Theorems 1 and 2. A bit more effort is needed to obtain Theorem 3 (Poisson distribution), but the result hardly differs from Theorem 2.
For multiple-quantum QR-PUF challenges, these are the first security proofs that do not require assumptions about the type of measurement performed by the adversary. It is interesting to note that the result for quadrature-based attacks [18] almost achieves equality in Theorem 2; quadrature measurements yield the optimal challenge estimation in practice.

Acknowledgments

We thank Berry Schoenmakers, Pepijn Pinkse and Allard Mosk for useful discussions, and anonymous reviewers for their constructive comments.

References

[1] R. Pappu, B. Recht, J. Taylor, and N. Gershenfeld. Physical One-Way Functions. *Science*, 297:2026–2030, 2002.

[2] B. Gassend, D.E. Clarke, M. van Dijk, and S. Devadas. Silicon Physical Unknown Functions. In *ACM Conference on Computer and Communications Security (CCS) 2002*, pages 148–160. ACM, 2002.

[3] J.D.R. Buchanan, R.P. Cowburn, A. Jausovec, D. Petit, P. Seem, G. Xiong, D. Atkinson, K. Fenton, D.A. Allwood, and M.T. Bryan. Forgery: 'fingerprinting' documents and packaging. *Nature, Brief Communications*, 436:475, 2005.

[4] P. Tuyls, G.J. Schrijen, B. Škorić, J. van Geloven, R. Verhaegh, and R. Wolters. Read-proof hardware from protective coatings. In *Cryptographic Hardware and Embedded Systems (CHES) 2006*, volume 4249 of *LNCS*, pages 369–383. Springer-Verlag, 2006.

[5] J. Guajardo, S.S. Kumar, G.J. Schrijen, and P. Tuyls. FPGA intrinsic PUFs and their use for IP protection. In *Cryptographic Hardware and Embedded Systems (CHES) 2007*, volume 4727 of *LNCS*, pages 63–80. Springer, 2007.

[6] G. DeJean and D. Kirovski. Radio frequency certificates of authenticity. In *IEEE Antenna and Propagation Symposium – URSI*, 2006.

[7] B. Škorić, T. Bel, A.H.M. Blom, B.R. de Jong, H. Kretschman, and A.J.M. Nellissen. Randomized resonators as uniquely identifiable anti-counterfeiting tags. *Secure Component and System Identification Workshop*, Berlin, March 2008.

[8] P. Tuyls, B. Škorić, and T. Kevenaar (Eds.). *Security with Noisy Data: Private Biometrics, Secure Key Storage and Anti-Counterfeiting*. Springer, London, 2007.

[9] A.-R. Sadeghi and D. Naccache (Eds.). *Towards Hardware-Intrinsic Security*. Springer, 2010.

[10] P. Tuyls, B. Škorić, S. Stallinga, A.H.M. Akkermans, and W. Ophey. Information-theoretic security analysis of Physical Uncloneable Functions. In *9th Conf. on Financial Cryptography and Data Security*, volume 3570 of *LNCS*, pages 141–155. Springer, 2005.

[11] T. Ignatenko, G.-J. Schrijen, B. Škorić, P. Tuyls, and F.M.J. Willems. Estimating the Secrecy Rate of Physical Uncloneable Functions with the Context-Tree Weighting Method. In *Proc. IEEE International Symposium on Information Theory (ISIT) 2006*, pages 499–503, 2006.

[12] B. Škorić. On the entropy of keys derived from laser speckle; statistical properties of Gabor-transformed speckle. *Journal of Optics A: Pure and Applied Optics*, 10(5):055304–055316, 2008.

[13] B. Škorić. Quantum Readout of Physical Unclonable Functions. *International Journal of Quantum Information*, 10(1):1250001–1 – 125001–31, 2012.

[14] B. Škorić. Quantum Readout of Physical Unclonable Functions. `http://eprint.iacr.org/2009/369`.

[15] S.A. Goorden, M. Horstmann, A.P. Mosk, B. Škorić, and P.W.H. Pinkse. Quantum-Secure Authentication with a Classical Key. `http://arxiv.org/abs/1303.0142`, 2013.

[16] W.K. Wootters and W.H. Zurek. A single quantum cannot be cloned. *Nature*, 299:802–803, 1982.

[17] D. Dieks. Communication by EPR devices. *Phys. Lett. A*, 92:271–272, 1982.

[18] B. Škorić, A.P. Mosk, and P.W.H. Pinkse. Security of Quantum-Readout PUFs against quadrature-based challenge-estimation attacks. *International Journal of Quantum Information*, 11(4):1350041–1 – 1350041–15, 2013.

[19] D. Bruß and C. Macchiavello. Optimal state estimation for d-dimensional quantum systems. *Phys. Lett. A*, 253(5-6):249–251, 1999.