Information-theoretic analysis of coating PUFs

B. Škorić, S. Maubach, T. Kevenaar, P. Tuyls
Philips Research Laboratories,
Prof. Holstlaan 6,
5656 AA Eindhoven, The Netherlands

Abstract *Physical Uncloneable Functions (PUFs) can be used as a cost-effective means to store cryptographic key material in an uncloneable way. In coating PUFs, keys are generated from capacitance measurements of a coating containing many randomly distributed particles with different dielectric constants.*

We introduce a physical model of coating PUFs by simplifying the capacitance sensors to a parallel plate geometry. We estimate the amount of information that can be extracted from the coating. We show that the inherent entropy is proportional to $\sqrt{n}(\log n)^{3/2}$, *where* n *is the number of particles that fit between the capacitor plates in a straight line. However, measurement noise may severely reduce the amount of information that can actually be extracted in practice. In the noisy regime the number of extractable bits is in fact a decreasing function of* n. *We derive an optimal value for* n *as a function of the noise amplitude, the PUF geometry and the dielectric constants.*

Keywords: Physical Uncloneable Function, PUF, Challenge-Response Pair, authentication, coating, dielectric, capacitance, entropy, key extraction

1 Introduction

1.1 General introduction to PUFs

A 'Physical Uncloneable Function' (PUF) is a function that is realized by a physical system, such that the function is easy to evaluate but the physical system is hard to characterize, model or reproduce.

Physical tokens were first used as identifiers in the 1980s in the context of strategic arms limitation treaty monitoring. The concept was later investigated for civilian purposes [1]. The tokens which were then studied are very hard to reproduce physically, but quite easy to read out completely, i.e. all the physical parameters necessary for successful identification are readily given up by the token. This makes these tokens suitable for systems where the verifier knows with certainty that an actual token is being probed and that the measuring device can be trusted. However, the tokens are not suitable for online identification protocols with a remote party. An imposter can relatively easily copy the data from someone's token, and then enter that data through a keyboard. The verifier cannot tell if a token is actually present.

Truly uncloneable tokens (PUFs) were introduced by Pappu [2, 3]. These tokens are so complex that it is infeasible to fully read out the data contained in a token or to make a computer model that predicts the outputs of a token [4]. This makes PUFs suitable for online protocols as well as verification involving physical probing by untrusted devices.

A PUF is a physical system designed such that it interacts in a complicated way with stimuli (*challenges*) and leads to unique but unpredictable *responses*. A PUF challenge and the corresponding response are together called a *Challenge-Reponse-Pair* (CRP). A PUF behaves like a keyed hash function; the physical system consisting of many 'random' components is equivalent to the key. In order to be hard to characterize, the system should not allow efficient extraction of the relevant properties of its interacting components by measurements. Physical systems that are produced by an uncontrolled production process, e.g. random mixing of several substances, turn out to be good candidates for PUFs. Because of this lack of control, it is hard to produce a physical copy of the PUF. Furthermore, if the physical function is based on many complex interactions, then mathematical modeling is also very hard. These two properties together are referred to as *Uncloneability*.

1.2 Applications

From a security perspective the uniqueness of the responses and uncloneability of the PUF are very useful properties. Because of these properties, PUFs can be used as unique identifiers [1, 5, 6, 7], means of tamper-detection and/or as a cost-effective source for key generation (common randomness) between two parties [8, 9]. By embedding a PUF inseparably into a device, the device becomes uniquely identifiable and uncloneable. Here 'inseparable' means that any attempt to remove the PUF will with very high probability damage the PUF and destroy the key material it contains. A wide range of devices can be equipped with a PUF in this way, e.g. smart-cards, credit cards, RFID tags, value papers, chips, security cameras, etc.

Several secure identification and authentication protocols based on CRPs have been worked out in [8, 10, 11]. Typically there are two phases: enrollment and verification. In the enrollment phase, a number of challenges is chosen randomly, and the corresponding PUF responses are measured and then stored in some form. In the verification phase the PUF is subjected to one or more of the enrollment challenges. The response is checked against the enrolled response data.

We distinguish between on the one hand 'identification', where a direct comparison is made between unprocessed PUF outputs, usually involving a correlation or distance measure, and on the other hand 'authentication', where a cryptographic key is derived from the PUF output for performing a cryptographic challenge-response protocol. In this paper we focus on the latter case. The typical scenario is that the verifier and the PUF holder are separated and communicate over an insecure channel.

For cryptographic protocols it is important to ensure that *exactly* the same bit string is derived from the enrollment and verification measurements in spite of the measurement noise. To this end so-called 'helper data' is generated for each CRP, data that describes how the PUF output should be processed, quantized etc. to obtain a noise-resilient bit string. The helper data for each enrolled challenge is stored together with the challenge. In most applications only the keys need to be kept secret. Hence, the challenges and helper data can be stored anywhere (e.g. conveniently on the PUF), while the keys must either be stored in a safe place or in some encrypted or hashed form. In the verification phase the verifier selects an enrolled challenge with the corresponding helper data. The PUF is subjected to this challenge and the PUF output is combined with the helper data to obtain a bit string. If this bit string

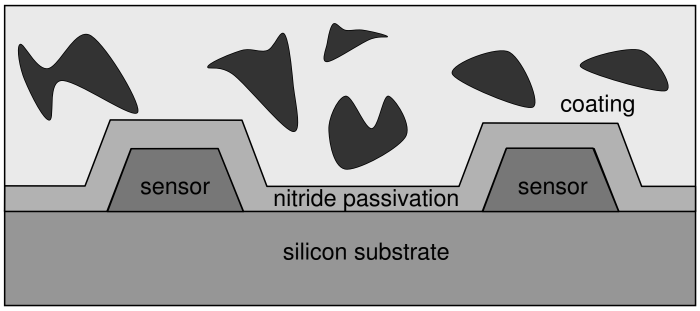

Figure 1: *Structure of a coating PUF. The sensor wires are perpendicular to the paper.*

is exactly equal to the enrolled key, then the cryptographic challenge-response protocol will result in a successful match, convincing the verifier that the PUF is authentic. Furthermore, at the end of the protocol the verifier and the PUF holder possess a shared secret that they can use e.g. as a session key. (Well designed protocols hide this key from eavesdroppers).

A special class of applications becomes possible if so-called 'control' is introduced [10]. A *Controlled PUF* (CPUF) is a PUF that is bound to a processor which completely governs the input and output. The chip can prohibit frequent challenging of the PUF and forbid certain classes of challenges. It can scramble incoming challenges. Furthermore, it can hide the physical output of the PUF, revealing to the outside world only indirect information derived from the output, e.g. an encryption or hash. This control layer substantially strengthens the security, since an attacker cannot probe the PUF at will and cannot interpret the responses. CPUFs allow for new applications such as 'certified execution' and 'certified measurement' [8, 10].

1.3 Coating PUFs

Several physical systems are known on which PUFs can be based. The main types are optical PUFs [2, 3], coating PUFs [8], silicon PUFs [11, 12] and acoustic PUFs [8]. In this paper we discuss coating PUFs. The idea of using an 'active coating' was proposed in [13] and further developed in the context of PUFs in [8].

Coating PUFs are integrated with an IC (see Fig. 1). The IC is covered with a coating consisting of e.g. aluminophosphate, which is doped with random dielectric particles. By random dielectric particles we mean several kinds of particles of random size and shape with a relative dielectric constant ε_r differing from the dielectric constant of the coating matrix. In order to challenge the coating PUF, an array of metal sensors (e.g. a comb structure of wires), is laid down directly beneath the passivation layer. Sufficient randomness is only obtained if the dielectric particles are approximately of the same size as the distance between the sensor parts, or smaller.

A challenge corresponds to a voltage of a certain frequency and amplitude applied to the sensors at a certain point of the sensor array. Because of the presence of the coating material with its random dielectric properties, the sensor plates with the material in between behave as a capacitor with a random capacitance value. The capacitance value is then converted into a bit string which can be used as an identifier or a key.

Coating PUFs have the advantage of possessing a high degree of integration. The matrix containing the random particles can be part of a tamper-resistance coating. A coating PUF also has the advantage that it is easily turned into a Controlled PUF (CPUF), as it is inseparably bound to the underlying device. The control electronics can simply be put underneath the coating.

1.4 Information-theoretical approach to PUFs

A general information-theoretical framework for the analysis of the security of PUFs was formulated in [4]. The central concept is the *entropy of a measurement*, i.e. the amount of information about the PUF's structure that is revealed by a measurement. One needs the notion of 'PUF space' or configuration space, a discrete space where each point corresponds to a possible PUF realisation. A measurement is represented as a partitioning of the PUF space, and the measurement entropy is the entropy of this partitioning. This formalism will be used for the analysis in sections 3 and 4.

1.5 Contributions of this paper

This paper contains the following novel contributions:

- We introduce a model of a coating PUF measurement at one location in the sensor array, by describing each sensor as a parallel plate capacitor. The geometry is simplified, but the effects of finite particle size are incorporated, as well as the insensitivity of the capacitance to particle permutations.

- Using our model, we compute the entropy of the probability distribution function of the capacitance. This 'inherent entropy' is the absolute upper bound on the extractable information. It corresponds to 'perfect' measurements, i.e. without any noise. The inherent entropy scales as $\sqrt{n}(\ln n)^{3/2}$, with n the number of particles that fits linearly between the capacitor plates.

- There are two counteracting effects at work. On the one hand, smaller particle size leads to more inherent PUF entropy. On the other hand, smaller particles imply better mixing, which leads to a reduced variance of the capacitance. (This is a 'law of large numbers' effect proportional to $1/\sqrt{\#\text{particles}}$). The latter puts a lower bound on the useful particle size, since a large capacitance variance is needed in order to obtain a good signal to noise ratio. We derive an optimum particle size that yields the highest number of extractable bits.

- In the regime of noisy measurements, the number of extractable bits is largest if (i) the relative dielectric constants of the two coating materials differ strongly, and (ii) the mixture contains only a small fraction of the substance with the low dielectric constant, namely of the order of the ratio of the two constants.

- If the measurement noise Δ is very small, of the order $1/(\text{density of states})$, individual capacitance states may be resolved. The density of states has a sharp peak, but the capacitances most likely to be measured lie outside this peak. Hence, if Δ is made so small that individual states can just be resolved, one enters into a regime where the finite density of states limits the extractable entropy, while the extractable entropy is still far smaller than the inherent entropy.

2 Modelling coating PUFs

2.1 Motivation

Our aim is to estimate the maximum amount of information that can be extracted from a coating PUF. To this end we formulate a physical model of a capacitance measurement (section 2.2) and compute the Shannon entropy of the capacitance distribution. We do not aim for an exact answer, but we want to know the order of magnitude and the scaling behaviour, i.e. the dependence of the entropy on all the important model parameters such as the distance between the sensor wires, the dielectric constants, the size of the random particles and the relative amounts of the random particles. We differentiate between two regimes:

1. **Measurements with very little noise.** In this case the amount of information that can be extracted is limited by the entropy of the PUF itself. The PUF entropy is finite due to the finite size of the random particles. The computation is presented in section 3.
2. **Noisy measurements.** In this case the finite particle size effects are unnoticeable, because they are overshadowed by the noise. The measurement entropy is completely determined by the signal to noise ratio. This computation is presented in section 4.

2.2 The model

For the sake of simplicity, we model the sensor wires and the coating above them as an ordinary capacitor consisting of two parallel electrode plates with a dielectric substance between them. This simplification will of course fail to represent the spatially varying electric field produced by the wires. However, we are interested only in the statistical properties of particle distributions within the region that contains most of the electric field density. As a first approximation, we idealize the geometry of the field.

As a first step we study the capacitor shown in Fig. 2a, a parallel-plate capacitor filled with layers $1 \dots n$ of equal thickness a/n with dielectric constants $\varepsilon_1 \dots \varepsilon_n$. It is well known [14] that its capacitance is given by

$$C_{n\text{ layers}} = C_{\text{ref}} \cdot \left(\frac{1}{n} \sum_{s=1}^{n} \frac{1}{\varepsilon_s} \right)^{-1} \quad ; \quad C_{\text{ref}} = \frac{A\varepsilon_0}{a}, \tag{1}$$

where A is the plate area, ε_0 the permittivity of the vacuum, and C_{ref} the capacitance of the system with vacuum between the plates, which we will use as a reference value throughout the paper. The result (1) has several invariance properties. A re-ordering of the layers does not change the capacitance. Additionally, C remains unchanged even if we split up a layer, so that we have more than n layers, and then re-order. In fact, as long as we make changes in the vertical direction only, the capacitance depends just on the average value of $1/\varepsilon$.

As a second step we look at the capacitor shown in Fig. 2b, with m columns of different dielectric material. This capacitor can, in good approximation, be considered as m parallel components, and hence its total capacitance is the sum of the parts,

$$C_{m\text{ columns}} = C_{\text{ref}} \cdot \frac{1}{m} \sum_{j=1}^{m} \varepsilon_j \tag{2}$$

We observe that only the average dielectric constant matters.

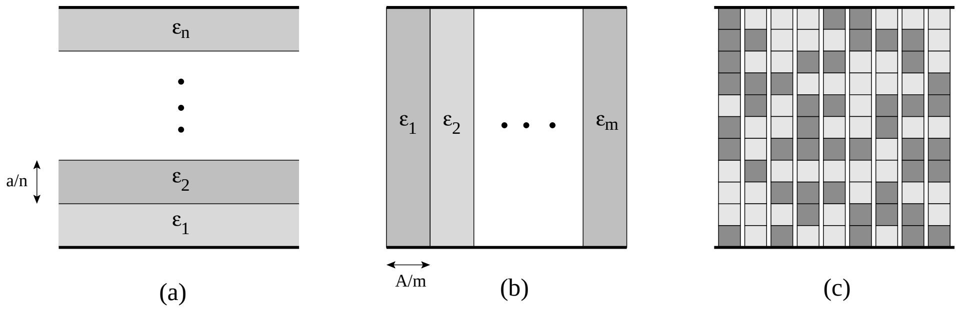

Figure 2: *Motivation of the model.* **(a)** *A capacitor consisting of several dielectric layers parallel to the plates.* **(b)** *Dielectric columns perpendicular to the plates.* **(c)** *Combination of layers and columns. The volume between the plates is filled with random dielectric building blocks.*

This leads us to the construction of our model (see Fig. 2c). Between the plates there is a mixture of two substances which have different dielectric constants, ε_1 and ε_2. Without loss of generality we will always assume that $\varepsilon_2 < \varepsilon_1$.

The volume is discretized: there are m columns of n 'voxels'. When the mixture is produced, the probability that a voxel will be occupied by substance 1 is denoted as p, and the probability of having substance 2 is $q = 1 - p$. The number of voxels in the j-th column that ends up filled with substance 1 is denoted as N_j. Writing the total capacitance as a sum of parallel column capacitances we have

$$C = \sum_{j=1}^{m} C_j \quad ; \qquad C_j = \frac{C_{\mathrm{ref}}}{m} \cdot n \left(\frac{N_j}{\varepsilon_1} + \frac{n - N_j}{\varepsilon_2} \right)^{-1} . \tag{3}$$

Note that C is invariant under swaps of complete columns and under voxel shifts within a column.

For convenience later on, we introduce the following notation. The number of columns containing precisely k particles of substance 1 ($k = 0 \dots n$) is denoted as α_k. The set $\{\alpha_k\}$ satisfies $\sum_{k=0}^{n} \alpha_k = m$, since the total number of columns is m. The capacitance is then expressed as

$$C = \sum_{k=0}^{n} \alpha_k \chi_k \quad ; \qquad \chi_k = \frac{C_{\mathrm{ref}}}{m} \cdot n \left(\frac{k}{\varepsilon_1} + \frac{n - k}{\varepsilon_2} \right)^{-1} . \tag{4}$$

Note that discrepancies may arise between our model and the geometry of Fig. 1 when the dielectric constants become very large. In our model the electric field lines are forced to move perpendicular to the plates, through the 'columns', while in Fig. 1 the field lines are free to avoid the coating altogether. However, we expect our model to be useful for reasonable values of $\varepsilon_1, \varepsilon_2$.

2.3 The density of states

First we examine the density of states (d.o.s.) in our model. The d.o.s. is the number of states that exist per infinitesimal interval on the capacitance axis, and we will denote it as

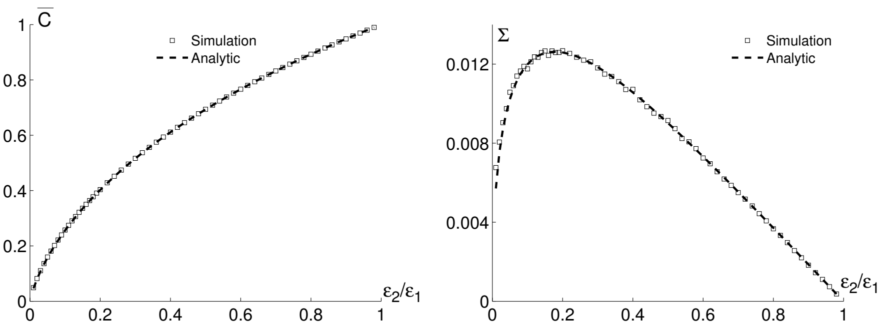

Figure 3: *Density of states for $n = 180$, $m = n$.* **Left:** *Location of the peak as a function of $\varepsilon_2/\varepsilon_1$.* **Right:** *Width of the peak as a function of $\varepsilon_2/\varepsilon_1$. The squares are numerical simulation results. The dashed curve in the left graph corresponds to (6). The dashed curve in the right graph is the estimate (7) with $\Gamma = 0.9$. All capacitances have been expressed in units of $C_{\text{ref}}\varepsilon_1$.*

$D(C)$. The total number of capacitance values in the model is given by the number of points in the α-lattice, i.e. the number of ways to partition m into $n+1$ non-negative integers, where the ordering is important.

$$N_{\text{states}} = \sum_{\alpha_0=0}^{m} \sum_{\alpha_1=0}^{m-\alpha_0} \cdots \sum_{\alpha_{n-1}=0}^{m-\alpha_0-\cdots-\alpha_{n-2}} 1 = \binom{n+m}{n}. \tag{5}$$

The d.o.s. must satisfy $\int_{C_{\text{ref}}\varepsilon_2}^{C_{\text{ref}}\varepsilon_1} D(C)\mathrm{d}C = N_{\text{states}}$. The states are distributed non-uniformly over the C-axis. In appendix B we estimate the shape of $D(C)$ based on a *typical set* argument. The highest concentration of states occurs at $C = C_{\text{peak}}$. For symmetry reasons, this point occurs when all the α_k are equal, i.e. $\alpha_k = m/(n+1)$ for all k. The corresponding capacitance C_{peak} is given by

$$C_{\text{peak}} \approx \frac{C_{\text{ref}}}{\varepsilon_2^{-1} - \varepsilon_1^{-1}} \ln \frac{\varepsilon_1}{\varepsilon_2}. \tag{6}$$

Terms of order $1/n$ are neglected. In the vicinity of this peak, the d.o.s. turns out to have an almost gaussian distribution with variance Σ,

$$\Sigma^2 \approx \Gamma^2 \frac{C_{\text{ref}}^2}{n} \cdot \left\{\varepsilon_1\varepsilon_2 - C_{\text{peak}}^2\right\}, \tag{7}$$

where Γ is a 'curve fitting' constant of order unity. Fig. 3 shows that (7) has good correspondence with simulation results. However, the typical set approximation is only valid close to the peak. The tails of the d.o.s. are not gaussian.

2.4 The probability distribution of the capacitance

Without loss of generality we assume that the ratio $\varepsilon_2/\varepsilon_1$ is chosen to be a non-algebraic number. In this way the mapping from $\{\alpha_k\}$ to C is bijective, i.e. the capacitance is uniquely determined by the set $\{\alpha_k\}$. (A proof is presented in Appendix A). This means that the probability distribution of C is equivalent to the probability distribution of $\{\alpha_k\}$. The latter is obtained as follows. First we introduce the notation x_k for the probability of finding k voxels with substance 1 in a given column. This is the binomial distribution,

$$x_k = \binom{n}{k} p^k q^{n-k}. \tag{8}$$

Then we note that the total probability of a configuration $\{\alpha_k\}$ is a multiplication of probabilities x_k, one for each column. Finally, the capacitance is invariant under column permutations, and hence the number of such permutations must be taken into account. This brings us to the following expression,

$$P_\alpha = \binom{m}{\alpha} \prod_{k=0}^{n} x_k^{\alpha_k} \quad ; \qquad \binom{m}{\alpha} = \frac{m!}{\alpha_0! \ \cdots \ \alpha_n!}. \tag{9}$$

Here we have used the shorthand notation $\alpha = \{\alpha_k\}$ with the implicit constraint $\sum_{k=0}^{n} \alpha_k = m$. It is easily verified that the probabilities P_α add up to unity using the following general identity [15],

$$\sum_\alpha \binom{m}{\alpha} \prod_{k=0}^{n} Y_k^{\alpha_k} = (Y_0 + \cdots + Y_n)^m. \tag{10}$$

A useful identity (for the computation of moments) can be derived from (10) by taking the derivative $\partial/\partial Y_s$,

$$\sum_\alpha \alpha_s \binom{m}{\alpha} \prod_{k=0}^{n} Y_k^{\alpha_k} = m Y_s (Y_0 + \cdots + Y_n)^{m-1}. \tag{11}$$

3 Entropy of a 'noiseless' measurement

3.1 Analytic part of the calculation

The goal is to compute the Shannon entropy H_α of the distribution (9). The first steps can be done analytically. We start by expanding $\ln P_\alpha$

$$H_\alpha = -\sum_\alpha P_\alpha \ln P_\alpha = -\sum_{k=0}^{n} \ln x_k \sum_\alpha P_\alpha \alpha_k - \sum_\alpha P_\alpha \ln \binom{m}{\alpha}. \tag{12}$$

The α-sum in the first right-hand-side term is evaluated using the identity (11) with $Y_k \to x_k$, yielding $\sum_\alpha P_\alpha \alpha_k = m x_k$. Rewriting the ln of the binomial in the last term as a sum of logarithms, we get

$$H_\alpha = -m \sum_{k=0}^{n} x_k \ln x_k - \ln m! + \sum_{k=0}^{n} \sum_\alpha P_\alpha \ln(\alpha_k!). \tag{13}$$

All three terms in (13) have a combinatorial interpretation. The first term is m times the entropy of the binomial distribution x_k. This represents the measurement entropy of m separate *distinguishable* columns. (Note that the capacitance measurement in our model does not 'see' the locations of columns.)

The second term is the entropy of permuting the m columns. The third term is the average entropy of permuting only those columns that have the same filling value k, for all k

separately. The second and third term together represent the average entropy of the $\binom{m}{\alpha}$ distinct column configurations that are consistent with a given set α.

The last term in (13) can be further evaluated. The α-sum averages a quantity that depends only on one component, α_k, of the set α. Hence the average w.r.t. the probability P_α can be replaced by the average w.r.t. the marginal distribution P_{α_k} of the component α_k.

$$H_\alpha = -m \sum_{k=0}^{n} x_k \ln x_k - \ln m! + \sum_{k=0}^{n} \sum_{\alpha_k=0}^{m} P_{\alpha_k} \ln(\alpha_k!). \tag{14}$$

The marginal distribution is given by

$$P_{\alpha_k} = \binom{m}{\alpha_k} x_k^{\alpha_k} (1-x_k)^{m-\alpha_k}. \tag{15}$$

The derivation is given in appendix C. Note that P_{α_k} is a binomial distribution corresponding to α_k out of m events with base probability x_k. This is what one would intuitively expect. As x_k itself is a binomial in k, we have 'nested' binomial distributions.

3.2 Approximation

We cannot evaluate the third term in (14) exactly. However, we can make a good approximation for $n \gg 1$, $m \gg \sqrt{n}$. (We remind the reader that $m \propto n$ in the 2D case and $m \propto n^2$ in the 3D case). We make use of the fact that both binomial distributions P_{α_k} and x_k are sharply peaked, and that x_k can be approximated by a normal distribution $\mathcal{N}_{np,\sigma}(k)$ in the vicinity of its peak, with $\sigma = \sqrt{npq}$. Furthermore, we define a constant c and an interval $I_c = (np - c\sigma, np + c\sigma)$ such that $mx_k > 1$ for $k \in I_c$. The details of the calculation are shown in appendix D. The result is

$$H_\alpha \approx \tfrac{1}{3} c^3 \sigma + \mathcal{O}(c\sigma) \quad ; \qquad c = f \sqrt{\ln \frac{m^2}{2\pi npq}}, \tag{16}$$

where f is a numerical constant of order one. Fig. 4 shows that the approximation is quite accurate.

There is an intuitive way of understanding the scaling $H_\alpha \propto c^3\sigma$. The entropy is approximately the log of the number of lattice points in the α-configuration lattice that carry substantial probability. The probability is concentrated around a sharp peak at $\langle \alpha_k \rangle = mx_k$. In each of the $n+1$ dimensions the standard deviation is $\sqrt{mx_k}$. However, in most of these dimensions $\sqrt{mx_k}$ is far less than one lattice point, and hence these hardly contribute to the entropy. In the contributing dimensions ($k \in I_c$) the standard deviation is of order $\sqrt{mx_{np}}$. Since $|I_c| = 2c\sigma$ we then get $H_\alpha \propto \ln(\sqrt{mx_{np}})^{2c\sigma} \approx c\sigma \ln(m/\sqrt{n}) \approx c^3\sigma$.

In the case of a two-dimensional capacitor, $m \propto n$, where the proportionality constant depends on the length and the width of the capacitor. In the three-dimensional case, m will scale as $m \propto n^2$. In both cases we have $\ln m \propto \ln n$, and therefore H_α scales as

$$H_\alpha \propto \sqrt{n} (\ln n)^{3/2}. \tag{17}$$

This equation for the entropy H_α is the main result of this section.

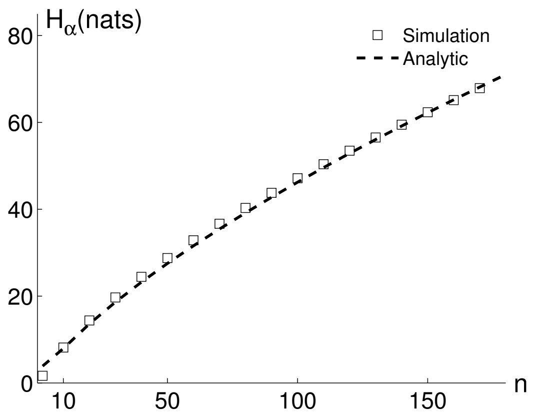

Figure 4: *Intrinsic entropy of a capacitor for $m = n, p = \frac{1}{2}$. The squares show the result of numerical evaluation of (14). The dotted curve is the approximation (50) with $f = 1.27$. The entropy is expressed in 'natural units', i.e. the logarithm with base e is used.*

4 Entropy of a noisy measurement

In the case of a noisy measurement, the noise is larger than the effects caused by the finiteness of the particle size. For all intents and purposes the capacitance can be treated as a continuous variable, i.e. a stochastic variable C with a smooth probability distribution function $\rho(C)$. In order to obtain reproducible measurements in spite of the noise, the C-axis is divided into bins of size Δ, where Δ is chosen proportional to the noise amplitude.

The entropy $H^\Delta[\rho]$ of the thus discretised distribution is given by [16]

$$H^\Delta[\rho] = h[\rho] - \ln \Delta, \tag{18}$$

where we have introduced the differential entropy $h[\rho] = -\int \mathrm{d}C \rho(C) \ln \rho(C)$. If the noise level is reduced, $h[\rho]$ remains constant, but the term $-\ln \Delta$ grows, and hence H^Δ grows. If the noise is made very small, the dielectric particle size becomes noticeable and (18) becomes invalid. Then one has to use the results of section 3.

The differential entropy $h[\rho]$ is readily estimated. In appendix E we give an approximation for the average μ_c and variance σ_c of the capacitance, using the model defined in section 2.2 and the capacitance distribution P_α (9). For $n \gg 1$ we have

$$\mu_c \approx \frac{C_{\mathrm{ref}}}{p\varepsilon_1^{-1} + q\varepsilon_2^{-1}} \quad ; \quad \sigma_c \approx \mu_c \sqrt{\frac{pq}{nm}} \cdot \frac{|\varepsilon_1^{-1} - \varepsilon_2^{-1}|}{p\varepsilon_1^{-1} + q\varepsilon_2^{-1}}. \tag{19}$$

Fig. 5 compares (19) to numerical simulations. The error in σ_c is 2%, while the error in μ_c is 0.2%.

Note that σ_c is a *decreasing* function of n and m. This can be understood as follows. When the number of random particles between the plates is large, the probability of deviating from the average value $\langle k \rangle = np$ is small for all columns. A finer mixing process allows for a better approximation of perfectly uniform mixing of the two substances.

If the capacitance distribution is sharply peaked ($\sigma_c/\mu_c \ll 1$) then we can replace it with a gaussian distribution without much loss of accuracy. The differential entropy of the gaussian

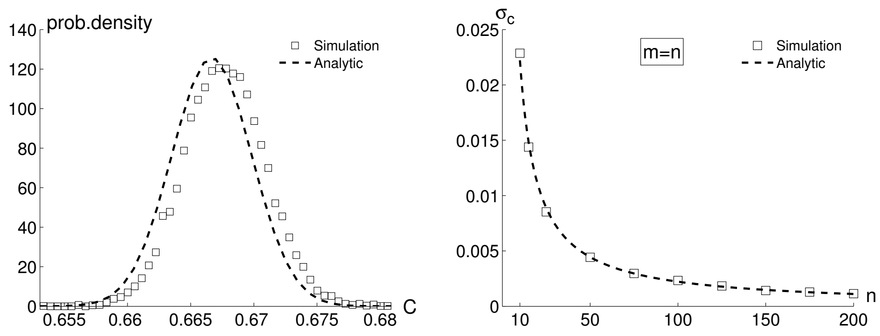

Figure 5: **Left:** *Probability distribution of the capacitance.* **Right:** *σ_c as a function of n. The dotted line represents the estimate (19). The squares show the statistical result of 10^4 randomly generated fillings. The parameters are $m = n = 70$, $p = \frac{1}{2}$, $\varepsilon_2 = \varepsilon_1/2$, and capacitances have been normalized w.r.t. $C_{\mathrm{ref}}\varepsilon_1$.*

distribution $\mathcal{N}_{\mu_c,\sigma_c}$ is given by [16]

$$h[\mathcal{N}_{\mu_c,\sigma_c}] = \ln(\sigma_c\sqrt{2\pi e}). \tag{20}$$

Combining (18) and (20) we can write the entropy of the discretised distribution as

$$H^{\Delta}[\rho] = \ln(\frac{\sigma_c}{\Delta}\sqrt{2\pi e}). \tag{21}$$

This equation has the form of a channel capacity for a noisy channel with signal to noise ratio $(\sigma_c/\Delta)^2$.

4.1 Optimal choice of $\varepsilon_1, \varepsilon_2, p$

In the derivation of (21) it was assumed that the distribution is sharply peaked. However, it is possible to obtain a rather broad distribution. Let $(1, \varepsilon_{\max})$, with $\varepsilon_{\max} \gg 1$, be the interval from which $\varepsilon_1, \varepsilon_2$ may be chosen. Take $\varepsilon_1 = \varepsilon_{\max}$ and make q and $\varepsilon_2/\varepsilon_1$ very small. In this limit we have

$$\sigma_c \to \frac{C_{\mathrm{ref}}\varepsilon_1}{\sqrt{nm}} \frac{\sqrt{q}\,\varepsilon_2/\varepsilon_1}{(q + \varepsilon_2/\varepsilon_1)^2}. \tag{22}$$

σ_c can be made large by choosing either (a) $q = \mathcal{O}(\varepsilon_2/\varepsilon_1) \ll 1$ or (b) $q \ll \varepsilon_2/\varepsilon_1 \ll \sqrt{q}$. For both approaches we show that the broadening of the distribution is not unlimited. The lowest feasible value of q must satisfy $q > 1/(nm)$. Otherwise, there would be only a small probability of having substance 2 in the capacitor at all, which clearly is not desirable.

- **Case (a):** We set $\varepsilon_2/\varepsilon_1 = \lambda q$, with λ a constant of order unity. This gives

$$\mu_c \to \frac{C_{\mathrm{ref}}\varepsilon_1}{1 + 1/\lambda} \quad ; \quad \sigma_c \to \frac{C_{\mathrm{ref}}\varepsilon_1\lambda}{(1+\lambda)^2}\frac{1}{\sqrt{qnm}} \leq \frac{C_{\mathrm{ref}}\varepsilon_1}{4\sqrt{qnm}} \tag{23}$$

  with equality for $\lambda = 1$. Since $q > 1/(nm)$, σ_c cannot exceed $C_{\mathrm{ref}}\varepsilon_1/4$, i.e. one quarter of the full capacitance range.

- **Case (b):** We can realize the choice $q \ll \varepsilon_2/\varepsilon_1 \ll \sqrt{q}$ by setting $\varepsilon_2/\varepsilon_1 = Bq^\gamma$, with B a constant of order unity and $\gamma \in (\frac{1}{2}, 1)$. This yields

$$\mu_c \to C_{\rm ref}\varepsilon_1 \quad ; \quad \sigma_c \to \frac{C_{\rm ref}\varepsilon_1}{B(nm)^{1-\gamma}}, \tag{24}$$

which is far smaller than $C_{\rm ref}\varepsilon_1$ since $nm \gg 1$.

We conclude that no reasonable choice of q and ε_2 can give rise to a σ_c exceeding $C_{\rm ref}\varepsilon_1/4$. In practice σ_c does not even come close to this value, because q has to be chosen much larger than $1/(nm)$ to get substantial mixing. Furthermore, in case a the parameter q cannot be much smaller than $1/\varepsilon_{\rm max}$, which may be a further impediment to lowering q.

It is also important to note that the intrinsic entropy H_α (16) becomes quite small in the limit of small q. In fact, a point q_0 may even exist where H_α gets smaller than $H^\Delta[\rho]$ (21). In that case (21) is clearly not the correct expression for the entropy. If we set $\varepsilon_2/\varepsilon_1 = q$ with $\frac{1}{nm} < q \ll 1$ (case a), then the crossover point $H^\Delta[\rho] = H_\alpha$ is given by the following implicit equation for q_0,

$$nq_0 \left[\ln \frac{m^2}{2\pi n q_0}\right]^3 = 9\left[\ln \frac{C_{\rm ref}\varepsilon_1\sqrt{2\pi e}}{4\Delta\sqrt{m}\sqrt{nq_0}}\right]^2 \tag{25}$$

Here we have put $f = 1$. Note that (25) can be read as a relationship between the three parameters m, $C_{\rm ref}\varepsilon_1/\Delta$ and nq_0. The extractable entropy in this regime is

$$H = \min\left\{\ln \frac{C_{\rm ref}\varepsilon_1\sqrt{2\pi e}}{4\Delta\sqrt{m}\sqrt{nq}},\ \frac{1}{3}\sqrt{nq}\left(\ln\frac{m^2}{2\pi nq}\right)^{3/2}\right\} \tag{26}$$

The point $\varepsilon_2/\varepsilon_1 = q = q_0$, with q_0 defined by (25), represents the optimal parameter choice yielding the highest possible entropy for fixed n and m. It is not a sharply peaked optimum, however, because of the weak q-dependence of $\ln(1/\sqrt{q})$ in $H^\Delta[\rho]$.

In this section we have always assumed that the measurement noise Δ is so large that many states fit inside a capacitance interval of width Δ, i.e. we have assumed $\Delta > 1/D(C)$. In the limit $\varepsilon_2/\varepsilon_1 \to 0$ the d.o.s. gets very sharply peaked around $C = C_{\rm ref}\varepsilon_2$, such that almost all states are concentrated there (see Fig. 3). This leaves few states in the vicinity of $C = \mu_c$. Thus, if Δ is small enough, $\Delta < 1/D(\mu_c)$ becomes a possibility. In that case (21,26) are valid no longer and the finite d.o.s. limits the extractable entropy.

5 Transition between noisy and noiseless regimes; Optimal n

In this section we investigate the limit of small Δ. As mentioned in the previous section, it can occur for small Δ that the extractable entropy is limited by the finite density of states. A transition between the 'noiseless' and 'noisy' regime takes place when the noise Δ is so small that individual states on the C-axis can be resolved. This happens when $1/\Delta$ is comparable in magnitude to $D(\mu_c)$. Taking this into account, (21) is replaced by

$$H^\Delta[\rho] = \ln\left(\sqrt{2\pi e}\min\left\{\frac{\sigma_c}{\Delta},\ D(\mu_c)\sigma_c\right\}\right), \tag{27}$$

where we have assumed that σ_c/μ_c is sufficiently small, so that the d.o.s. in the interval $(\mu_c - \sigma_c, \mu_c + \sigma_c)$ is approximately constant at $D(\mu_c)$. (We are not looking at 'case a' here).

Unfortunately, the results of section 2.3 do not give us the d.o.s. at $C = \mu_c$, since in general μ_c does not lie close to $C_{\rm peak}$ (6). We have to resort to another type of approximation to determine the d.o.s. in the tail of the $D(C)$ distribution. Note that, since μ_c lies in this tail, the transition effects are noticeable, long before each individual state on the whole C-axis can be resolved! Consequently, the extractable entropy (27) will be significantly smaller than the intrinsic entropy H_α even when $D(\mu_c) < 1/\Delta$.

We define δC as the smallest capacitance step that we can generate by applying (integer) changes $\delta\alpha_k$ to the 'average' configuration $\alpha_k = mx_k$. This gives us the estimate $D(C) \approx 1/\delta C$. The best method we could identify is to center the $\delta\alpha_k$ parameters around the center of the α_k distribution ($k = np$), and to arrange them in such a way that they generate an N-th derivative. For instance, if we take

$$\delta\alpha_{np+v} = (-1)^{v+1} \binom{N}{\frac{1}{2}N + v}, \quad v \in \{-\tfrac{1}{2}N, \ldots, \tfrac{1}{2}N\} \tag{28}$$

then the capacitance step $\delta C = \sum_k \chi_k \delta\alpha_k$ is the discretised derivative $\frac{\partial^N}{\partial k^N}\chi_k$ at $k = np$. It turns out that for $n > 10^5$ the best result is obtained at $N = 2$. For $n < 10^5$ the optimal N can be much higher. However, the allowed values of $\delta\alpha_k$ are bounded by the 'starting values' mx_k, and this bounds N. The highest allowed N satisfies $\delta\alpha_{np} = -mx_{np}$ with $\delta\alpha_{np}$ defined in (28). Using Stirling's approximation for the binomial, the bound can be expressed as

$$\frac{2^{N+1}}{\sqrt{N}} = \frac{m}{\sqrt{npq}}. \tag{29}$$

The resulting δC is

$$\delta C \approx \left. \frac{\partial^N}{\partial k^N}\chi_k \right|_{k=np} = C_{\rm ref}\frac{N!}{mn^N}\,\frac{(\varepsilon_2^{-1} - \varepsilon_1^{-1})^N}{(p\varepsilon_1^{-1} + q\varepsilon_2^{-1})^{N+1}}. \tag{30}$$

When δC (30) is equal to Δ, then the transition between the 'noisy' and 'noiseless' regime takes place.

The point of equality, $\Delta = \delta C$, can be seen as an equation expressing the transition value of Δ as a function of n, m, ε_1, ε_2 and p. Conversely, the equation can also be read to give the transition values for n, m as a function of Δ, ε_1, ε_2 and p. If we write $m = \lambda n^{d-1}$, with λ a proportionality constant and d the number of dimensions (2 or 3) of the capacitor, then the transition value of n is given by the implicit equation

$$n_{\rm trans} = \left[\frac{C_{\rm ref}N!(\varepsilon_2^{-1} - \varepsilon_1^{-1})^N}{\lambda\Delta(p\varepsilon_1^{-1} + q\varepsilon_2^{-1})^{N+1}}\right]^{\frac{1}{N+d-1}} \tag{31}$$

where N depends on $n_{\rm trans}$ logarithmically according to (29),

$$N \approx \log_2 \frac{\lambda}{e\sqrt{pq}} + (d - \tfrac{3}{2}) \log_2 n_{\rm trans}. \tag{32}$$

Eq. (31) roughly defines the optimal particle size for a given noise level. On the one hand, larger particles lead to a smaller number of distinguishable capacitance values (δC grows) within the region of high probability, and hence the measurable entropy decreases. On the other hand, taking smaller particles also reduces the measurable entropy, since the ratio σ_c/Δ

decreases. (Remember that $\sigma_c \propto 1/\sqrt{n}$). For small changes in n, N is almost a constant, and (31) gives a power law dependence $n_{\rm trans} \propto (1/\Delta)^{\frac{1}{N+d-1}}$, with logarithmic corrections. Using this approximation, we see that the full Δ-dependence of the entropy H^Δ (following from $\sigma_c \propto (nm)^{-\frac{1}{2}}$) is given by $-\frac{N-1+d/2}{N-1+d} \ln \Delta$.

6 Summary

Coating PUFs are a cost-effective way of storing cryptographic key material in an uncloneable way. An IC is covered with a coating that is doped with random dielectric particles. A secret bit string is derived from capacitance measurements.

We have introduced a simplified physical model of coating PUFs by representing each sensor in the PUF as a parallel plate capacitor. Using this model, we have computed the intrinsic entropy of the mixture between the capacitor plates as a function of the relative amounts of the two substances, the number of 'columns' (m), and the number of 'slots' in each column (n). For large n and m, the intrinsic entropy scales as $\sqrt{n}(\ln[m/\sqrt{n}])^{3/2}$.

The actually extractable information can be significantly lower due to measurement noise. The entropy of a capacitance measurement is dictated by the signal to noise ratio σ_c/Δ, where the 'signal' is the variance σ_c of the capacitance distribution and the noise Δ is the uncertainty in the measured capacitance. The variance scales as $1/\sqrt{nm}$, reflecting the fact that large deviations from average filling become increasingly unlikely when the mixing becomes finer. For fixed number of particles, a large variance σ_c is obtained if one of the dielectric constants is very large ($\varepsilon_1 \gg 1$) and the other close to 1, while the mixing is such that the ε_1 material is far more abundant than the other ($q = \mathcal{O}(\varepsilon_2/\varepsilon_1)$). However, σ_c cannot be made arbitrarily large without a penalty, since (a) the inherent entropy, which scales as $\sqrt{q}$, will become too small, and (b) the mixing ratio q has to be larger than $1/(nm)$ in order for the ε_2 material to be present at all.

At fixed ε_1, ε_2, p, the extractable entropy has a maximum as a function of the particle size when the discrete capacitance steps in the model are approximately equal to the noise Δ. The optimal particle size scales as a power of Δ, with logarithmic corrections.

ACKNOWLEDGEMENTS
We thank Geert-Jan Schrijen, Rob Wolters, Nynke Verhaegh, Jan van Geloven, Arnold Gruijthuijsen and Hennie Kretschman for useful discussions.

References

[1] D.W. Bauder, *An Anti-Counterfeiting Concept for Currency*, Systems Research Report PTK-11990, Sandia National Laboratories, 1983.

[2] R. Pappu, *Physical One-Way Functions*, Ph.D. thesis, MIT 2001.

[3] R. Pappu, B. Recht, J. Taylor, N. Gershenfeld, *Physical One-Way Functions*, Science Vol. 297, Sept 2002, p.2026.

[4] P. Tuyls, B. Škorić, S. Stallinga, A.H.M. Akkermans, W. Ophey, *Information-Theoretic Security Analysis of Physical Uncloneable Functions*, A.S. Patrick and M.

Yung (Eds.): Proc. 9th Conf. on Financial Cryptography and Data Security , March 2005, LNCS 3570, pp. 141-155, `http://fc05.ifca.ai/p12.pdf`

[5] Unicate BV's '3DAS' system, `http://www.andreae.com/Unicate/Appendix%201.htm`, 1999.

[6] D. Kirovski, *A Point-Subset Compression Algorithm for Fiber-based Certificates of Authenticity*, IEEE Proc. ISIT 2004, p.173.

[7] J.D.R Buchanan, R.P. Cowburn, A. Jausovec, D. Petit, P. Seem, G. Xiong, D. Atkinson, K. Fenton, D.A. Allwood, M.T. Bryan, *Forgery: 'Fingerprinting' documents and packaging* , Nature 436, p.475 (28 Jul 2005), Brief Communications

[8] P. Tuyls, B. Škorić, *Secret Key Generation from Classical Physics*, Proc. Hardware Technology Drivers for Ambient Intelligence Symposium, Philips Research Book Series, Kluwer, 2005.

[9] B. Škorić, P. Tuyls, W. Ophey, *Robust key extraction from Physical Uncloneable Functions*, Ioannidis, Keromytis, Yung (Eds.): Proc. ACNS 2005, LNCS 3531, pp.407-422.

[10] B. Gassend, D. Clarke, M. van Dijk, S. Devadas, *Controlled Physical Random Functions*, Proc. 18th Annual Computer Security Applications Conf., Dec. 2002.

[11] B. Gassend, *Physical Random Functions*, Master's Thesis, MIT 2003.

[12] B. Gassend, D. Clarke, M. van Dijk, S. Devadas, *Silicon Physical Random Functions*, Proc. 9th ACM Conf. on Computer and Communications Security, Nov. 2002.

[13] R. Posch, *Protecting Devices by Active Coating*, Journal of Universal Computer Science, vol.4, no.7 (1998), pp.652-668.

[14] See e.g. D.J. Griffiths, *Introduction to electrodynamics*, Prentice-Hall, 1981.

[15] See e.g. S. Ross, *A first course in probability* (6th ed.), p.247, Prentice-Hall, 2001.

[16] See e.g. T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley series in telecommunications, 1991.

[17] P. Flajolet, *Singularity analysis and asymptotics of Bernoulli sums*, Theoretical Computer Science 215(1-2), pp.371–381, 1999.

A Bijective mapping $\alpha \leftrightarrow C$

In this appendix we prove that choosing a non-algebraic value for $\varepsilon_2/\varepsilon_1$ implies a bijective mapping between α and C. (A number is called non-algebraic if it cannot be represented as the solution of a polynomial equation with integer coefficients).

Let's assume that two vectors $\alpha^{(1)}$ and $\alpha^{(2)}$ yield the same value C. Then $\sum_k d_k \chi_k = 0$, where we have defined $d_k = \alpha_k^{(1)} - \alpha_k^{(2)}$. We rewrite this equation as $\sum_k d_k/(k\varepsilon_2/\varepsilon_1 + n - k) = 0$.

Multiplying by $\prod_j (j\varepsilon_2/\varepsilon_1 + n - j)$ we get

$$\sum_{k=0}^{n} d_k \prod_{j|j\neq k} (j\frac{\varepsilon_2}{\varepsilon_1} + n - j) = 0. \tag{33}$$

This is a polynomial equation in $\varepsilon_2/\varepsilon_1$ with integer coefficients. Since $\varepsilon_2/\varepsilon_1$ is not an algebraic number, the equation can only be satisfied if $d_k = 0$ for all k. Hence $\alpha^{(1)} = \alpha^{(2)}$, which completes the proof.

B Density of states; Typical set

In this appendix we estimate the shape of the $D(C)$ function. To this end we treat the α_k as stochastic variables with a uniform distribution between 0 and m, subject to the collective constraint $\sum_{k=0}^{n} \alpha_k = m$. In other words, we employ a 'fake' uniform distribution instead of the actual non-uniform probability distribution P_α (9). Consequently, all points in the α-lattice are treated equally. This construction allows us to determine the d.o.s. numerically by Monte Carlo simulation.

We use the concept of the 'typical set' [16] of α-configurations. When drawing a random α from the uniform distribution, there is an overwhelming probability that it will belong to the typical set.

First we determine the 'mean' capacitance C_{peak}. For symmetry reasons, this point occurs when all the α_k are equal, i.e. $\alpha_k = m/(n+1)$ for all k. The corresponding capacitance C_{peak} is given by

$$C_{\text{peak}} = \sum_{k=0}^{n} \frac{m}{n+1}\chi_k \approx \int_0^1 \mathrm{d}\beta \, \frac{C_{\text{ref}}}{\beta\varepsilon_1^{-1} + (1-\beta)\varepsilon_2^{-1}} = \frac{C_{\text{ref}}}{\varepsilon_2^{-1} - \varepsilon_1^{-1}} \ln\frac{\varepsilon_1}{\varepsilon_2}. \tag{34}$$

Here we have used the definition of χ_k (4) and we have approximated the sum by an integral by introducing $\beta = k/n$ (i.e. $\sum_k \to n\int \mathrm{d}\beta$). Furthermore we have neglected terms of order $1/n$.

We determine the shape of the d.o.s. curve by taking the continuum approximation, i.e. we treat the α_k as continuous variables on the interval $[0, m]$. The number of states that satisfy the two constraints $\sum_k \alpha_k = m$ and $\sum_k \alpha_k \chi_k = C$ can be expressed as an integral over two Dirac delta functions that enforce those constraints,

$$D(C) \propto \int_0^m \mathrm{d}\alpha_0 \cdots \int_0^m \mathrm{d}\alpha_n \; \delta(\sum_{k=0}^{n} \alpha_k - m)\delta(\sum_{k=0}^{n} \alpha_k\chi_k - C). \tag{35}$$

We perform a basis transformation $\alpha_k \to \eta_k$ that simplifies the first delta function. We define $\eta_0 = (n+1)^{-1/2}\sum_{k=0}^{n}\alpha_k$ and $\eta_j = (\alpha_0 - \alpha_j)/\sqrt{2}$ for $1 \le j \le n$. The inverse relations are $\alpha_0 = \eta_0/\sqrt{n+1} + \sqrt{2}/(n+1)\sum_{j=1}^{n}\eta_j$ and $\alpha_k = \eta_0/\sqrt{n+1} + \sqrt{2}/(n+1)\sum_{j=1}^{n}\eta_j - \sqrt{2}\eta_k$ for $1 \le k \le n$. In the new basis the integrals are of the form

$$\begin{aligned} D(C) \quad \propto & \quad \int \mathrm{d}\eta_0 \cdots \int \mathrm{d}\eta_n \; \delta(\eta_0 - \frac{m}{\sqrt{n+1}}) \\ & \quad \delta(C_{\text{peak}}\frac{\sqrt{n+1}}{m}\eta_0 + \frac{\sqrt{2}}{m}\sum_{j=1}^{n}\eta_j[C_{\text{peak}} - m\chi_j] - C). \end{aligned} \tag{36}$$

The integration intervals of the η variables are more complicated than in (35). We Integrate out the first delta function, which leads to the replacement $\eta_0 \to m/\sqrt{n+1}$. Next we use an integral representation for the second delta function according to $\delta(x) = (2\pi)^{-1} \int_{-\infty}^{\infty} \mathrm{d}p\; e^{ipx}$. The exp of the sum nicely factors into a product where each factor only depends on a single η_j variable.

$$D(C) \propto \int_{-\infty}^{\infty} \frac{\mathrm{d}p}{2\pi} e^{-ip[C-C_{\text{peak}}]} \int \mathrm{d}\eta_1 \cdots \int \mathrm{d}\eta_n \prod_{j=1}^{n} \exp ip\frac{\sqrt{2}}{m}\eta_j [C_{\text{peak}} - m\chi_j]. \tag{37}$$

However, the factorisation is incomplete in the sense that the $\eta_1 \cdots \eta_n$ integrals cannot be evaluated independently, as the integration bounds on each η variable are affected by the other η variables.

At this point we introduce an approximation: We estimate the integration intervals based on the properties of the typical set. First of all, from the symmetry between the α_k it follows that the bounds on all the α_j do not depend on j. Furthermore, we can think of the d.o.s. as a probability distribution for C based on continuous variables α_k, such that each point in α-space is equally likely. In this view, α_k does not deviate much from its 'average' value $m/(n+1)$ in the set of typical configurations. Since α_k has to stay nonnegative, the magnitude of this deviation will be of the order $m/(n+1)$. Recalling the definition $\eta_j = (\alpha_0 - \alpha_j)/\sqrt{2}$, we take an estimated interval $\eta_j \in [-m\Gamma/(n+1)\sqrt{2}, +m\Gamma/(n+1)\sqrt{2}]$. Here we have introduced a numerical constant Γ of order unity which reflects our ignorance. Note that our approximation is valid only in the vicinity of $C = C_{\text{peak}}$, i.e. inside or close to the typical set.

Each η-integral is evaluated independently, and the result is

$$\begin{aligned} D(C) &= N_{\text{states}} \int_{-\infty}^{\infty} \mathrm{d}p\; e^{-ip(C-C_{\text{peak}})} G(p) \\ G(p) &\approx \prod_{k=1}^{n} \operatorname{sinc} \frac{p\Gamma(m\chi_k - C_{\text{peak}})}{n+1}, \end{aligned} \tag{38}$$

where 'sinc' denotes the function $\operatorname{sinc} x = x^{-1} \sin x$. The $G(p)$ is the generating function for the distribution of the variable $C - C_{\text{peak}}$. All moments of this distribution can be obtained by differentiating G at $p = 0$. As $G(p)$ is even in p, it is clear that all odd moments are zero. The width Σ of the distribution is given by

$$\Sigma^2 = - \left.\frac{\partial^2 G}{\partial p^2}\right|_{p=0} = \frac{\Gamma^2}{(n+1)^2} \sum_{k=1}^{n} (m\chi_k - C_{\text{peak}})^2. \tag{39}$$

The summation of χ_k^2 can be approximated by an integration as before, with $\beta = k/n$,

$$m^2 \sum_{k=0}^{n} \chi_k^2 \approx n \int_0^1 \mathrm{d}\beta\; [\beta\varepsilon_1^{-1} + (1-\beta)\varepsilon_2^{-1}]^{-2} = n\varepsilon_1\varepsilon_2, \tag{40}$$

yielding the result (7)

$$\Sigma^2 \approx \Gamma^2 \frac{C_{\text{ref}}^2}{n} \cdot \left\{\varepsilon_1\varepsilon_2 - C_{\text{peak}}^2\right\}. \tag{41}$$

C Marginal distribution of α_k

In this appendix we determine the marginal probability distribution (15) of α_k. The computation goes as follows. We start with the distribution (9) for the whole set α. One variable α_k is singled out of the α-summation, leaving all $\{\alpha_j\}$ with $j \neq k$. Then the identity (10) is used to evaluate the summation over these n variables.

For some arbitrary function f we can write

$$\sum_{\alpha} P_{\alpha} f(\alpha_k) = \sum_{\alpha_k=0}^{m} \binom{m}{\alpha_k} x_k^{\alpha_k} f(\alpha_k) \sum_{\{\alpha_j\}, j\neq k} \frac{(m-\alpha_k)!}{\prod_{t\neq k} \alpha_t!} \prod_{s,s\neq k}^{n} x_s^{\alpha_s}. \tag{42}$$

The identity (10), but now for the variables $\alpha \backslash \alpha_k$, gives

$$\sum_{\alpha} P_{\alpha} f(\alpha_k) = \sum_{\alpha_k=0}^{m} \binom{m}{\alpha_k} x_k^{\alpha_k} (1-x_k)^{m-\alpha_k} f(\alpha_k) =: \sum_{\alpha_k=0}^{m} P_{\alpha_k} f(\alpha_k). \tag{43}$$

D Approximate entropy

In this appendix we approximate the summations in the last term of (14). The computation consists of three steps.

1. We note that the binomial distribution P_{α_k} is sharply peaked around $\langle \alpha_k \rangle = m x_k$. The value $m x_k$ is vanishingly small when k lies in one of the tails of the binomial distribution x_k. Hence, for 'tail' values of k, the contribution to the k-sum will be approximately $P(\alpha_k = 0)\ln(0!) + P(\alpha_k = 1)\ln(1!) + P(\alpha_k = 2)\ln(2!) = P(\alpha_k = 2)\ln(2)$, which is vanishingly small because of the negligible $P(\alpha_k = 2)$.

   This means that we only have to sum over those values k that lie in the peak of the distribution x_k. The peak is centered on $k = np$ and has standard deviation $\sigma = \sqrt{npq}$. Our summation interval is $I_c = [np - c\sigma, np + c\sigma]$, where the constant c is somewhat arbitrarily defined such that at the boundaries $m x_k \approx 1$. We have

$$m x_{np \pm c\sigma} \approx 1 \quad ; \qquad c^2 = 2f^2 \ln \frac{m}{\sqrt{2\pi npq}}, \tag{44}$$

   where f is a numerical constant of order unity. Because of the somewhat fuzzy definition of c, we have to 'fit' f to obtain the correct proportionality constant in H_α. It turns out that the best choice for f has a weak dependence on p and lies between 1.1 and 1.3.

   Note that we need $m > \sqrt{n}$, otherwise a solution does not exist.

2. For $k \in I_c$ the distribution P_{α_k} is sharply peaked around $\langle \alpha_k \rangle > 1$. In general, for a sharp distribution of some variable u and some smooth function $f(u)$ one can approximate $\langle f(u) \rangle \approx f(\langle u \rangle)$. Using this technique, we can write the third term in (14) as

$$\begin{aligned} &\sum_{k=0}^{n} \sum_{\alpha_k=0}^{m} P_{\alpha_k} \ln(\alpha_k!) \approx \sum_{k\in I_c} \ln(m x_k)! \\ &\approx \sum_{k\in I_c} \ln \sqrt{2\pi m x_k} + m \sum_{k\in I_c} x_k \ln x_k + m \ln \tfrac{m}{e} \sum_{k\in I_c} x_k. \end{aligned} \tag{45}$$

In the last step we have used the Stirling approximation. Substitution of (45) into (14) gives

$$H_\alpha \approx \sum_{k\in I_c} \ln\sqrt{2\pi m x_k} - m\ln\tfrac{m}{e}\sum_{k\in\text{tail}} x_k - m\sum_{k\in\text{tail}} x_k \ln x_k, \tag{46}$$

where we have neglected terms of order $\ln m$.

3. We can approximately evaluate the summations in (46) by replacing x_k for $k\in I_c$ by a gaussian distribution with average np and standard deviation $\sigma=\sqrt{npq}$. The approximation holds for $|k-np|\ll n$.

   Then we replace the summations by integrations. The first term gives

$$\int_{np-c\sigma}^{np+c\sigma} \mathrm{d}k \ln\sqrt{2\pi m x_k} \approx \tfrac{1}{3}c^3\sigma + c\sigma\ln\sqrt{2\pi}. \tag{47}$$

   In the second term of (46) we get

$$1-\int_{np-c\sigma}^{np+c\sigma}\mathrm{d}k\; x_k \approx 1-\mathrm{Erf}\tfrac{c}{\sqrt{2}} \approx \sqrt{\tfrac{2}{\pi}}c^{-1}e^{-c^2/2} = \frac{2\sigma}{mc}, \tag{48}$$

   where we have used the asymptotic expansion of the Erf function for large arguments.

   For the computation of the third term in (46) we note that the full k-sum would yield the entropy H_{np} of the binomial distribution, which we know [17] to be $H_{np}=\ln(\sigma\sqrt{2\pi e})+\mathcal{O}(1/n)$. We calculate the tail entropy as the full entropy H_{np} minus the entropy in the peak,

$$\begin{aligned} H_{np}+\int_{np-c\sigma}^{np+c\sigma}\mathrm{d}k\; x_k\ln x_k &\approx [1-\mathrm{Erf}\tfrac{c}{\sqrt{2}}]\ln(\sigma\sqrt{2\pi e}) - \tfrac{c}{\sqrt{2\pi}}e^{-c^2/2} \\ &\approx \frac{2\sigma}{mc}\ln(m\sqrt{e}). \end{aligned} \tag{49}$$

   Comparison of (47), (48) and (49) shows that the first term in (46) has contributions of order $c^3\sigma$ and $c\sigma$, while the second and third term in (46) partly cancel each other, leaving only a contribution of order c/σ.

The final result is

$$H_\alpha \approx \tfrac{1}{3}c^3\sigma + c\sigma\ln\sqrt{2\pi} + 3\sigma/c. \tag{50}$$

E Average and variance of the capacitance

In this appendix we estimate the average μ_c and standard deviation σ_c of the capacity $C=\sum_k \alpha_k\chi_k$ (4). We have $\mu_c=\langle C\rangle_\alpha=\sum_k\chi_k\langle\alpha_k\rangle_\alpha$ and

$$\sigma_c^2=\langle C^2\rangle_\alpha-\langle C\rangle_\alpha^2=\sum_{k,l=0}^{n}\chi_k\chi_l[\langle\alpha_k\alpha_l\rangle_\alpha-\langle\alpha_k\rangle_\alpha\langle\alpha_l\rangle_\alpha] \tag{51}$$

where the notation $\langle\cdot\rangle_\alpha$ indicates averaging with respect to P_α (9). We use the identity (11) to compute the expectation values analytically,

$$\langle\alpha_k\rangle_\alpha = m x_k \quad ; \quad \langle\alpha_k\alpha_l\rangle_\alpha-\langle\alpha_k\rangle_\alpha\langle\alpha_l\rangle_\alpha = -m x_k x_l + \delta_{kl} m x_k. \tag{52}$$

Substitution into (51) gives

$$\mu_c = m \langle \chi_k \rangle \quad ; \quad \sigma_c^2 = m \left[\left\langle \chi_k^2 \right\rangle - \langle \chi_k \rangle^2 \right]. \tag{53}$$

The notation $\langle \cdot \rangle$ indicates averaging w.r.t. the distribution x_k. Analytic computation does not yield a closed-form solution. Hence we approximate as follows. We define $k = np + \sigma u$, where $\sigma = \sqrt{npq}$ is the standard deviation of x_k. In terms of the new variable u, which is of order 1 in the peak of x_k, we can write

$$\chi_k = \frac{C_{\mathrm{ref}}}{m[p\varepsilon_1^{-1} + q\varepsilon_2^{-1}]} \cdot \frac{1}{1 + u\psi} \quad ; \qquad \psi = \frac{\sigma}{n} \cdot \frac{\varepsilon_1^{-1} - \varepsilon_2^{-1}}{p\varepsilon_1^{-1} + q\varepsilon_2^{-1}}. \tag{54}$$

For $n \gg 1$ we can make use of the fact that $\psi = \mathcal{O}(1/\sqrt{n})$ to make a Taylor expansion in ψ to second order. Since $\langle u \rangle = 0$ and $\left\langle u^2 \right\rangle = 1$ we obtain

$$\begin{aligned} &\left\langle \frac{1}{1+\psi u} \right\rangle \approx 1 + \psi^2 \\ &\left\langle \frac{1}{(1+\psi u)^2} \right\rangle - \left\langle \frac{1}{1+\psi u} \right\rangle^2 \approx (1 + 3\psi^2) - (1 + 2\psi^2) = \psi^2. \end{aligned} \tag{55}$$

This results in

$$\mu_c = \frac{C_{\mathrm{ref}}}{p\varepsilon_1^{-1} + q\varepsilon_2^{-1}} \quad ; \quad \sigma_c = C_{\mathrm{ref}} \cdot \sqrt{\frac{pq}{nm}} \cdot \frac{\left|\varepsilon_1^{-1} - \varepsilon_2^{-1}\right|}{[p\varepsilon_1^{-1} + q\varepsilon_2^{-1}]^2}. \tag{56}$$