# Permutation-Based Hash from Non-Idealized Assumptions: Adding Feed-Forward to Sponge*

Chun Guo[123], Kai Hu[12367], Shuntian Jiang[123],
Yanhong Fan[123], Yong Fu[4], Bart Preneel[5], and Meiqin Wang[1236(✉)]

[1] School of Cyber Science and Technology, Shandong University, Qingdao, China
[2] State Key Laboratory of Cryptography and Digital Economy Security, Shandong University, Qingdao, 266237, China
[3] Key Laboratory of Cryptologic Technology and Information Security of Ministry of Education, Shandong University, Qingdao, Shandong, 266237, China
[4] Qilu University of Technology, Jinan 250100, Shandong, China
[5] imec-COSIC, KU Leuven, Leuven, Belgium
[6] Quan Cheng Shandong Laboratory, Jinan, China
[7] Suzhou Research Institute, Shandong University, Suzhou, 215123, China chun.guo.sc@gmail.com,{kai.hu, yanhongfan,202517221,yongfu,mqwang}@sdu.edu.cn,bart.preneel@esat.kuleuven.be

**Abstract.** Understanding the security of permutation-based algorithms has been highlighted (by e.g., Naya-Plasencia and Daemen) as a central open question in symmetric cryptography. While convincing treatments have been proposed for keyed schemes, the landscape for permutation-based (keyless) hash functions remains bleak. We study permutation-based (keyless) hash functions using Rogaway's Human Ignorance approach. We show that by adding feed-forward back to (the inner part of the state of) the sponge hash construction, the collision and (second) preimage security of the resulting construction Sponge-F can be reduced to *simple and well-defined properties* of the underlying permutation, i.e., correlation intractability w.r.t. a certain class of relations. The reductions hold in the quantum setting as well, enabling proving the first meaningful quantum bounds for permutation-based hash constructions using Cojocaru et al.'s lifting theorem. As a bonus, in the random permutation model, the non-quantum (second) preimage security bounds of Sponge-F are much better than the standard sponge with the same capacity.

**Keywords:** Hash function · sponge construction · human ignorance

# Table of Contents

## 1 Introduction

Permutation-based cryptography has become increasingly popular. A cryptographic permutation is a keyless public permutation that is designed to behave as a random permutation. Compared to a blockcipher, a permutation avoids the costs and potential weakness of key schedule functions. The use of cryptographic permutation gained popularity during the SHA-3 competition, where several candidates were based on permutations, and the permutation-based KECCAK sponge function became the final winner [37]. Soon after, permutation-based PRNGs [14], MACs [4,34], and (authenticated) encryption [15] were proposed. Notably, during the NIST lightweight competition [74], 6 out of the 10 finalists are permutation-based. This success has also motivated the community to initiate a dedicated workshop for relevant advances and questions [85].

The security of these schemes is usually proved in the *random permutation model (RPM)*, that is, the permutation is randomly chosen, and all parties are given (black-box) access to it. RPM suffers from two caveats. First, in theory, it is well-known that no permutation with a short description is random [23,18]. Moreover, one cannot justify how "well" a permutation (such as the KECCAK-P) "behaves as a random permutation", since the latter is not a well-defined property. Second, for efficiency, permutation designers tend to live with full-round distinguishers. For example, the KECCAK-P 1600-bit permutation admits full 24-round zero-sum distinguishers with $2^{1573}$ complexity. However, instead of *increasing* the number of rounds to eliminate zero-sum distinguishers, the designers tend to further *reduce* rounds (and have proposed new algorithms using 12-round KECCAK-P [89]). The hopes are two-fold: (i) Zero-sum distinguishers do not yield non-trivial collision nor preimage attacks against the hash functions. (ii) Relevant security claims typically only withstand attacks with complexity below $2^{512}$, which is well below the cost of a zero-sum distinguisher. *However, there is no rigorous theoretical model that could formally prove these intuitions, at least for VIL hash functions (see below).*

Due to these, *understanding security from keyless permutations* has been recognized as one of the most important next steps in symmetric cryptography, and has been highlighted in Naya-Plasencia's Eurocrypt 2022 invited talk [71], in the summary of 2022 Dagstuhl workshop Symmetric Cryptography [56] as one of the four most important topics, and in Daemen's book chapter [33]). In particular, [56] challenged understanding *what non-random properties of permutations seem likely to be translated into an attack on the hash.* In other words, the following open question can be formulated:

*Which properties of permutations are sufficient to prove the security of the cryptographic hash function?*

Successful attempts have been made regarding *keyed constructions*: keyed sponge security has been reduced to the security of the underlying Even-Mansour cipher [4,82,34]; universal hash and MAC security have been reduced to differential properties [41]; AEAD and tweakable enciphering scheme security have been reduced to security of permutation-based deck ciphers [46,9]. But these are inapplicable to keyless hash functions. For the latter, one may consider the RPM with differential-style weaknesses [67], or the *public-seed pseudorandom permutation* assumption [82]: but the former remains an ideal model, while the latter remains inapplicable to keyless hash functions.

For keyless hash functions, only obvious reductions [27,83] are known for very simple fixed-input-length (FIL) constructions, whereas the landscape of VIL hash remains wide open. Recall that the most popular VIL permutation-based hash construction is the sponge construction of Bertoni et al. [12,45]. At a high level, the sponge operates on a state of size $b$ bits, which is split into an inner part of size $c$ bits (the capacity) and an outer part of size $r$ bits (the rate), where $b = c + r$. The sponge consists of an absorbing phase and a squeezing phase. In the absorbing phase, data is compressed into the state $r$ bits at a time, interleaved with an evaluation of a $b$-bit permutation $\mathsf{P}$. In the squeezing phase, a digest is extracted from the state $r'$ bits at a time (possibly $r' \neq r$, as proposed by Guo et al. [47]), again interleaved with an evaluation of $\mathsf{P}$. The whole construction is depicted in Fig. 1. The sponge found quick adoption right after its introduction, and featured in both the NIST SHA-3 competition winner KECCAK [37] and the NIST LWC competition winner ASCON-HASH256 [84].

An attractive feature of the sponge is avoiding the memory cost of feed-forward operations. But this means sponge is somewhat invertible: given a $b$-bit state, one can invoke $\mathsf{P}$ and $\mathsf{P}^{-1}$ to evaluate both forwardly and backwardly in the construction. This invertibility makes it harder to answer our main question. Concretely, since the permutation is keyless, it is natural to apply Rogaway's Human Ignorance approach [79] to the sponge. We have attempted this, but were only able to reduce sponge hash security to intricate properties over *multiple correlated permutation inputs/outputs*, which resemble the "multi-block CICO problem" [45, Sect. 8.2.5] (see App. A). The reason is that the multiple permutation-calls appeared during a sponge hash computation are "closely linked" by the invertibility, and cannot be treated separately.

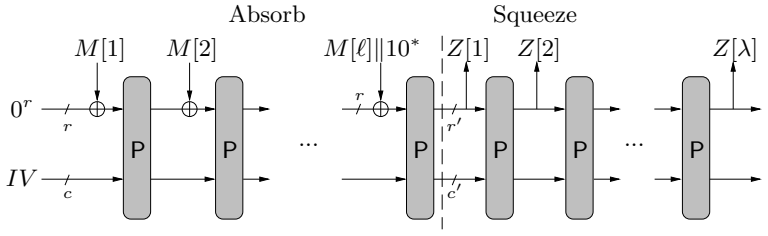


**Fig. 1.** The standard (generalized) sponge construction: the absorbing phase uses rate $r$ and capacity $c$, while the squeezing phase uses rate $r'$ and capacity $c'$. The final output is of $\lambda r'$ bits $Z[1]\|...\|Z[\lambda]$.

### 1.1 Our Contributions: Sponge-with-feed-forward with Security from Non-idealized Assumptions

Recently, Foekens [40] studied sponge variants using a *non-invertible* random function, and proved that non-invertibility increases (second) preimage security to optimal $\min\{c, h\}$ bits. Another natural idea to break invertibility is to *add feed-forward back* to the permutation-based sponge. While this seems a step "backward" (since the aforementioned feature of low memory cost is lost), we investigate this idea and give results as follows.

- We propose a concrete construction $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ with output length $h$ and capacity $c$, which: (i) feeds the $c$-bit inner part of the input forward to the output in the sponge absorbing phase; (ii) squeezes outputs from the *inner part* of the state in squeezing phase. The construction is shown in Fig. 2.
- When $h \leq c$, we prove that in both classical and quantum settings, the collision and (second) preimage security can be reduced to *non-idealized assumptions*, i.e., correlation intractability (CI) of the underlying permutation w.r.t. a specific binary relation and some unary relations. To formalize CI for keyless permutations, we follow Rogaway's Human Ignorance approach [79] and serve some initial discussions. These solve our main question.
- By studying the relations and applying Cojocaru et al.'s lifting theorem [28] (which lifts results in the classical random permutation model to the quantum random permutation model), we establish concrete security bounds for $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ (when $h \leq c$) in the random permutation model:
  - The classical collision, everywhere preimage and everywhere second preimage security are $h/2$-bit, $h$-bit and $\min\{h, c - \log_2 \ell\}$-bit respectively, where $\ell$ is the number of blocks in the second preimage challenge message. The (second) preimage bounds are much better than the standard sponge with the same parameters $h$ and $c$, due to which a smaller permutation could be used and the cost of the partial feed-forward can be compensated for.
  - For quantum collision, everywhere preimage and everywhere second preimage, we establish $h/4$, $h/2$ and $\min\left\{h/2, \frac{c-\log_2 \ell}{2}\right\}$-bit security respectively. To our knowledge, this is the first meaningful quantum security bounds for permutation-based hash constructions.
- As application, we show that Ascon-p-based $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ instantiation can be used in NIST LMS signature [29] to reduce its hardware area.

Below we elaborate in detail.

**Our Construction $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$.** Our hash construction is illustrated in Fig. 2. Concretely, for each permutation call we *feed the $c$-bit inner part* of the input forward to the output (intuitively, the outer part can be controlled by adversary-chosen message blocks, and feeding it forward does not appear helpful). In the squeezing phase, the output is squeezed from the *inner part* of the state, at the rate of $r'$ bits each time, as illustrated in Fig. 2 (Top). As shown by Andreeva et al. [6], outputting the inner part may enable length-extension attacks. To preclude these, when the last padded message block $\overline{M}[\ell]$ is XORed to the outer part of the state, a non-zero (but fixed) constant $\theta \in \{0,1\}^c$ is simultaneously XORed to the inner part. This borrows ideas from the Merkle-Damgård with permutation construction of Hirose et al. [48].

We require that $r' \leq c$. When used as a fixed-output-length hash function with $h$-bit output, we additionally require that $h \leq r' \leq c$—but in this case, one can simply set $h = r' = c$ and output the $c = h$ bit inner part of the final state, as shown in Fig. 2 (Bottom). We call this *full-inner-squeezing*.

By adding feed-forward and outputting the inner part of the state, the FOL variant of $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ becomes quite close to a Merkle-Damgård construction using the truncated Davies-Meyer construction $\mathsf{TrDM}^{\mathsf{P}}_c(X) :=$

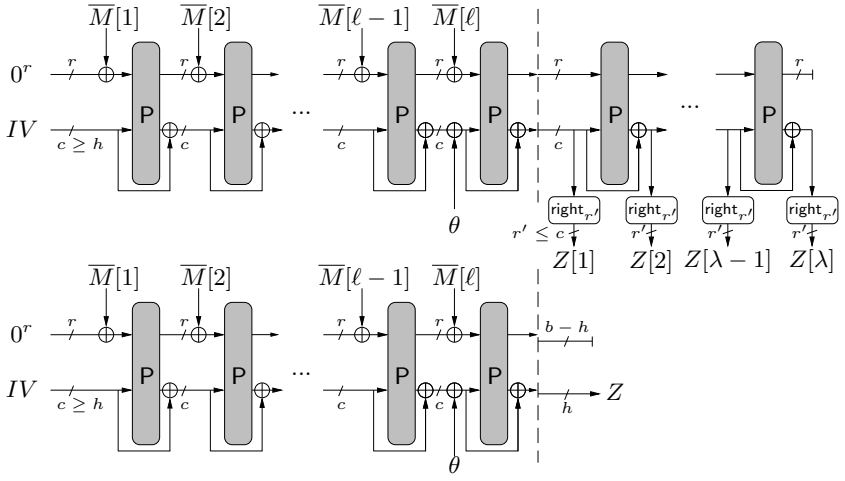


**Fig. 2.** (Top) Construction $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ from sponge-with-feed-forward. (Bottom) With fixed output length $h \leq c$, one can choose $r' = h$ to have a construction with $h/2$-bit collision and $h$-bit preimage security.

$\mathsf{right}_c\big(X \oplus \mathsf{P}(X)\big)$ as compression function ($\mathsf{right}_c(X)$ returns the rightmost $c$ bits of $X$). We will serve a more detailed comparison in Sect. 1.3.

**Reducing Security to Correlation Intractability (CI).** When $h \leq c$, the similarity between $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ and Merkle-Damgård construction using the truncated Davies-Meyer $\mathsf{TrDM}_c^{\mathsf{P}}$ enables us to reduce collision and preimage security of $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ to collision and preimage security of $\mathsf{TrDM}_c^{\mathsf{P}}$, which is then virtually equivalent with *correlation intractability (CI) of the permutation* $\mathsf{P}$ *w.r.t. certain (family of) relations*. This notion means that no efficient adversary can find inputs and outputs of the permutation $\mathsf{P}$ to satisfy the given relations. But we need to overcome subtle technical issues.

(i) First, the notion of CI was introduced by Canetti et al. [23] for *keyed functions*, whereas we consider *keyless* hash and permutations. The gap is not merely syntactical. For example, consider the binary relation $R_{\text{coll}}$, where $\big((X, X'), (Y, Y')\big) \in R_{\text{coll}}$ iff. $\mathsf{right}_c(X \oplus Y) = \mathsf{right}_c(X' \oplus Y')$: one cannot assume that no efficient adversary can find $\big((X, X'), (\mathsf{P}(X), \mathsf{P}(X'))\big) \in R_{\text{coll}}$ for a keyless permutation $\mathsf{P}$, since an adversary (in the non-uniform model) can simply embed such a pair $(X, X')$ in its code. This resembles the *keyless dilemma* for collision resistant hash.
To overcome this, we adapt Rogaway's Human Ignorance approach [79] to CI. Roughly speaking, instead of assuming that no adversary can find $\big((X, X'), (\mathsf{P}(X), \mathsf{P}(X'))\big) \in R_{\text{coll}}$ for $\mathsf{P}$, we explicitly construct such a CI adversary $\mathscr{B}$ from a collision adversary $\mathscr{A}$ against $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$.

(ii) Second, everywhere preimage security is meaningless for keyless functions, even in Human Ignorance setting (we refer to Sect. 2.1 for details), and we need to consider alternatives. For this, we use the notion *Z-preimage security*, which quantifies the hardness of finding preimage $(\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}})^{-1}(Z)$ for the specific challenge image $Z$. This idea was also used by Stinson [86, Sect. 4]. As its security assumption, we consider an unary relation $R_{\text{pre}}(Z)$, where $(X, Y) \in R_{\text{pre}}(Z)$ iff. $\mathsf{right}_{|Z|}(X \oplus Y) = Z$.

(iii) Third, second preimage security is related to the unary relation $R_{\text{spr}}(X, \nu)$, where $(X', Y') \in R_{\text{spr}}(X, \nu)$ iff. $\mathsf{right}_\nu(X' \oplus Y') = \mathsf{right}_\nu\big(X \oplus \mathsf{P}(X)\big)$ and $\nu \in \{c, h\}$. But everywhere second preimage security is meaningless for keyless functions either. If we consider second preimage security for each particular challenge message $M$, then we can only show (in Lemma 3) that *each challenge $M$ can be compiled into a sequence of $b$-bit "challenge" strings $X_1, ..., X_\ell$, such that finding second preimage for $M$ indicates finding $\big(X', \mathsf{P}(X')\big)\big) \in R_{\text{spr}}(X_i, c)$ for some $X_i \in \{X_1, ..., X_\ell\}$*. But it is not clear what this means for a $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$.
To reach a "more meaningful" conclusion, we consider *distribution-oriented* second preimage security notion of [7], which requires the challenge message $M$ to be sampled according to a distribution $\mathbf{d}$.

With the above, collision and (second) preimage security of $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ are (roughly) reduced to CI of $\mathsf{P}$ w.r.t. the binary relation $R_{\text{coll}}$ and the unary relations $R_{\text{pre}}(Z)$ and $R_{\text{spr}}(X, \nu)$:

(i) Given an adversary finding a collision of $\mathsf{Sponge}\text{-}\mathsf{F}^{\mathsf{P},\mathsf{pd}}$, we can build either an adversary finding $\big((X, X'), (\mathsf{P}(X), \mathsf{P}(X'))\big) \in R_{\mathrm{coll}}$, or an adversary finding $(X, \mathsf{P}(X)) \in R_{\mathrm{pre}}(IV)$. We remark that for efficiency concern, $\mathsf{Sponge}\text{-}\mathsf{F}^{\mathsf{P},\mathsf{pd}}$ does not use suffix-free padding, and the relation $R_{\mathrm{pre}}(IV)$ has to be used;

(ii) Given an adversary finding a preimage $(\mathsf{Sponge}\text{-}\mathsf{F}^{\mathsf{P},\mathsf{pd}})^{-1}(Z)$ for the challenge image $Z$, we can build an adversary finding $(X, \mathsf{P}(X)) \in R_{\mathrm{pre}}(Z)$;

(iii) Given a samplable distribution $\mathbf{d}$ on the message space and an adversary finding the second preimage $\mathsf{Sponge}\text{-}\mathsf{F}^{\mathsf{P},\mathsf{pd}}(M') = \mathsf{Sponge}\text{-}\mathsf{F}^{\mathsf{P},\mathsf{pd}}(M)$ for $M$ sampled according to $\mathbf{d}$, one of the following can be built:
   - An adversary finding $\big((X, X'), (\mathsf{P}(X), \mathsf{P}(X'))\big) \in R_{\mathrm{coll}}$ using the sampler of $\mathbf{d}$;
   - An adversary finding $(X, \mathsf{P}(X)) \in R_{\mathrm{pre}}(IV)$;
   - Another samplable distribution $\mathbf{d}'$ that samples a sequence of $b$-bit strings $X_1, ..., X_\ell$, $\ell$ equals the number of blocks in $M$ after padding, as well as an adversary that can find $(X, \mathsf{P}(X)) \in R_{\mathrm{spr}}(X_i, c)$ for one of them.

   Namely, the distribution-oriented second preimage security of $\mathsf{Sponge}\text{-}\mathsf{F}^{\mathsf{P},\mathsf{pd}}$ is reduced to CI of $\mathsf{P}$ w.r.t. $R_{\mathrm{coll}}$, $R_{\mathrm{pre}}(IV)$ and a *distribution-oriented multi-target CI notion*, in which the challenge relations are sampled from the family $\{R_{\mathrm{spr}}(X, c), X \in \{0,1\}^b\}$. We refer to Theorem 3 for formal elaborations, as well as its interpretation.

While the ideas are simple, this seems the first time permutation-based VIL hash functions have their security based on well-defined properties on underlying permutations, and also the first treatment of CI in a keyless setting.

*On validating our assumptions.* It is much easier to estimate the security margins of standard permutations (such as Keccak-p or Ascon-p) regarding our assumptions of CI w.r.t. the relations $R_{\mathrm{coll}}$, $R_{\mathrm{pre}}(Z)$ and $R_{\mathrm{spr}}(X, \nu)$, than to "directly" estimate the collision and preimage security of VIL hash constructions, and the results could support *using reduced-round permutations*. Since Keccak-p and Ascon-p have not been officially used in truncated Davies-Meyer mode, we are not aware of any cryptanalysis of Keccak-p/Ascon-p w.r.t. our CI properties. We made preliminary analyses of the (preimage-related) unary relation $R_{\mathrm{pre}}(Z)$, with a focus on differential-linear (or 1-bit truncated differential) cryptanalysis, which appears the most effective statistical cryptanalytic method on Keccak-p and Ascon-p. We followed a state-of-the-art preimage attack idea of Niu et al. [75] with the assistance of a search tool of Hu et al. [51], with results as follows:

- For Ascon-p, we did not find differential-linear distinguishers against 7-round Ascon-p with correlation $> 2^{-64}$ (6-round Ascon-p does admit a distinguisher with a single-bit output mask and correlation $2^{-62}$). Such a low correlation has been harmless enough [75], meaning that for CI w.r.t. $R_{\mathrm{pre}}(Z)$, full 12-round Ascon-p has a security margin of 5 rounds against differential-linear attacks;
- For Keccak-p, 4 rounds are free of useful differential-linear bias, so 12 rounds of Keccak-p already provide a high security margin for us.

While this paper certainly could not provide a complete analysis of the properties against all known attack methods (and our analyses are very preliminary), the above does demonstrate the feasibility.

In the past decade, the community has emphasized *efficiency* over *security margin*. For example, the Keccak designers proposed to reduce the number of rounds of Keccak-p from 24 to 12 for fast hashing [16], and the resulting algorithms have been published as RFC 9861 [89]. At Real World Crypto 2020, Aumasson [8] also advocated reducing rounds in primitives, and proposed to use 10-round Keccak-p (which is even less than [16]). Reducing rounds in primitives should be supported by thorough analyses of their security margins, to which our construction and approach could provide support.

We remark that our design choices are crucial for the above reductions and preliminary estimation. First, the added feed-forward "isolates" the computations between different permutation invocations and enables the reductions to treat them separately, which in turn enables using CI w.r.t. simple unary or binary relations as the underlying assumptions. This eases cryptanalytic testifications. Second, due to the idea of full-inner-squeezing, the output of $\mathsf{Sponge}\text{-}\mathsf{F}^{\mathsf{P},\mathsf{pd}}$ is the output of the final $\mathsf{TrDM}^{\mathsf{P}}_h$ invocation, and this enables unifying the reduction assumptions.

In comparison, reduction proofs for the standard sponge have to rely on much more complicated relations of polynomial arity (we refer to App. A), and cryptanalytic validations basically have to consider the sponge function "as a whole". This increases the difficulty in estimating security margins.

**Concrete bounds in (quantum) random permutation model.** By further studying hardness of the aforementioned relations in the random permutation model, we are able to derive concrete bounds: for fixed

$h$-bit output Sponge-$\mathsf{F}^{\mathsf{P},\mathsf{pd}}$ with $c \geq h$, the collision, everywhere preimage and everywhere second preimage security of $h$-bit output Sponge-$\mathsf{F}^{\mathsf{P},\mathsf{pd}}$ are roughly of $h/2$, $h$ and $\min\{h, c - \log_2 \ell\}$ bits (since we are now in the random permutation model, everywhere-style notions become meaningful), where $\ell$ is the number of blocks in the second preimage challenge message. The collision bound is comparable with the standard sponge $\min\left\{2^{c/2}, 2^{h/2}\right\}$, while the preimage and second preimage bounds are much better than that of the standard sponge, which are $\min\left\{\max\left\{2^{h-r'}, 2^{c/2}\right\}, 2^h\right\}$ and $\min\left\{2^{c/2}, 2^h\right\}$ respectively [5,59].

Moreover, using the quantum lifting theorem of [28], our results can be extended to the quantum random permutation model, yielding *quantum* collision, everywhere preimage and everywhere second preimage security of roughly $h/4$, $h/2$ and $\min\left\{h/2, \frac{c-\log_2 \ell}{2}\right\}$ bits, where $\ell$ is the number of blocks in the second preimage challenge message. The (second) preimage bounds are tight, while collision is not. As will be discussed later, this appears the first meaningful quantum bounds for variable-input-length (VIL) permutation-based hash constructions.

*Indifferentiability.* To complement, in Sect. 6 we also prove that Sponge-$\mathsf{P}^{\mathsf{P},\mathsf{pd}}$ is indifferentiable from a variable-output-length (VOL) random oracle, up to approximately $\min\{2^{c/2}, 2^{b-r'}\}$ queries. The squeeze rate $r'$ thus limits indifferentiability security, meaning that $r'$ cannot be too large when Sponge-$\mathsf{F}^{\mathsf{P},\mathsf{pd}}$ is used as an indifferentiable XOF. (We do not claim this as our main contribution.)

### 1.2 Instantiations and Applications

Our construction and bounds can be used in two directions:

- First, fixing a permutation, our construction enables achieving higher (second) preimage security. Concretely, using the NIST standard permutations Ascon-p [84] and Keccak-p [37], we propose: (i) Ascon-Sp-F, an Ascon-p-based FOL instantiation of Sponge-$\mathsf{F}^{\mathsf{P},\mathsf{pd}}$ with better (second) preimage security than standard Ascon-Hash256; (ii) Keccak-Sp-F$_{768}$, a Keccak-p-based FOL instantiation of Sponge-$\mathsf{F}^{\mathsf{P},\mathsf{pd}}$ with better (second) preimage security than SHA-3;
- Second, fixing concrete collision and (second) preimage security bits as the goal, our constructions enable either using a larger rate (and thus higher throughput) or using a smaller permutation (and thus smaller hardware area). We propose Keccak-Sp-F$_{512}$ that achieves the same level of collision and (second) preimage security as SHA-3-512 with throughput roughly 1.6 times that of SHA-3-512.

Below, we elaborate on the instances in detail.

Ascon-Sp-F*: Enriching* Ascon *family and reducing area of LMS signature* Using the NIST 320-bit permutation Ascon-p [84], we instantiate FOL construction Sponge-$\mathsf{F}^{\mathsf{P},\mathsf{pd}}$ (Fig. 12 Bottom) with $h = c = 256$ and propose Ascon-Sp-F. We compare its state size with Ascon (and a concurrent design Ascon-DM [88]). With fewer fed-forward bits and single-call squeezing, Ascon-Sp-F is both smaller and faster than Ascon-DM. Compared with Ascon-Hash256, despite the increased hardware area, Ascon-Sp-F achieves better throughput for short messages with less than 64 bytes (e.g., in LMS signature that will be discussed). Meanwhile, Ascon-Sp-F has better (second) preimage security and provable quantum security. Table 1 serves comparisons among the security bits. Ascon-Sp-F thus may be a useful complement to the Ascon family.

**Table 1.** Comparison between Ascon-Sp-F, Ascon-DM of Sun et al. [88] and Ascon-Hash256. All parameters are expressed in bits. There is no quantum security proof for Ascon-DM. For (quantum) second preimage security (q-)spre., security is considered for challenge messages with up to $2^{64}$ blocks (after padding). The algorithm Ascon-EDM is obtained by instantiating the SPONGE-EDM$^{\mathsf{P},\mathsf{pd}}$ construction of [88] with Ascon-p, but [88, Sect. 1.4] explicitly stated that Ascon-EDM is not suggested (we thus mark it in gray).

| Algorithm | State | $h$ | $r$ | $c$ | Col. | Pre. | spre. | q-col. | q-pre. | q-spre. | Ref. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Ascon-Hash256 | 320 | 256 | 64 | 256 | 128 | 192 | 128 | 26.7 | 26.7 | 26.7 | [84] |
| Ascon-DM | 640 | 256 | 64 | 256 | 128 | 192 | 192 | - | - | - | [88] |
| Ascon-EDM | 576 | 256 | 64 | 256 | 128 | 192 | 192 | - | - | - | [87] |
| Ascon-Sp-F | 576 | 256 | 64 | 256 | 128 | 256 | 192 | 64 | 128 | 96 | Ours |
| SHA-3-512 | 1600 | 512 | 576 | 1024 | 256 | 512 | 512 | 133.3 | 133.3 | 133.3 | [84] |
| Keccak-DM[576, 512] | 3200 | 512 | 1024 | 576 | 256 | 512 | 512 | - | - | - | [88] |
| Keccak-EDM$^c$[576, 512] | 2176 | 512 | 1024 | 576 | 256 | 512 | 512 | - | - | - | [88] |
| Keccak-Sp-F$_{512}$ | 2176 | 512 | 1024 | 576 | 256 | 512 | 512 | 128 | 256 | 256 | Ours |
| Keccak-DM[1088, 1024] | 3200 | 1024 | 512 | 1088 | 512 | 1024 | 1024 | - | - | - | [88] |
| Keccak-EDM$^c$[1088, 1024] | 2688 | 1024 | 512 | 1088 | 512 | 1024 | 1024 | - | - | - | [88] |
| Keccak-Sp-F$_{1024}$ | 2688 | 1024 | 512 | 1088 | 512 | 1024 | 1024 | 256 | 512 | 512 | Ours |

The improved (second) preimage security motivated applying our results to hash-based signatures. We consider the Leighton-Micali Signature (LMS) [61], which has been published as RFC 8554 [64] and standardized as RFC 9708 [50], 9802 [42] and NIST SP 800-208 [29]. Its state-of-the-art security proof was due to Fluhrer [39], who formalized a complicated "LMS-(specific-)multi-target second preimage security" for hash functions and proved that SHA-256 satisfies this definition. We adapt Fluhrer's proof [39] to prove that Ascon-Sp-F could replace SHA-256 in LMS: Ascon-Sp-F ensures an upper bound of $2^{-60}$ on the success probability of attacks with complexity $2^{120}$.

Compared with SHA-256 and SHAKE, Ascon-Sp-F has a much smaller state (Ascon-Sp-F: $\approx 256 + 320 = 576$ bits; SHA-256: $\approx 256 + 512 = 768$ bits; SHAKE: 1600 bits), and is more suitable for constrained devices. The standard Ascon-Hash256 may also be used, but (at present) its security can only be derived from indifferentiability [57], which only ensures $2^{-16}$ success probability for attacks with $2^{120}$ complexity. As discussed in [64, Sect. 9.1], a much lower success probability bound is indeed desirable for such standards (NIST requires $2^{-32}$ probability [44,36]). Ascon-Sp-F is also more efficient than Ascon-Hash256: the majority of hash evaluations in LMS have 54 or 55 bytes inputs, on which Ascon-Sp-F performs better (as discussed).

*For Chinese standardization:* Keccak-Sp-$\mathrm{F}_{1024}$. In 2025, China announced calls for new generation cryptographic algorithms, in which a hash function supporting $2^{64}$-bit messages and achieving 1024-bit digests and 1024-bit preimage and second preimage security is required [73]. Such a high level of security cannot be achieved by the standard 1600-bit permutation Keccak-p via existing constructions (this includes the recent construction of Lefevre and Mennink [60] with $2c/3$-bit indifferentiability, since $2c/3 = 1024$ means the construction can only have a 64-bit rate). Our results provide a perfect solution for this requirement: instantiating FOL construction $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ with $h = 1024$ and $c = 1088$ yields Keccak-Sp-$\mathrm{F}_{1024}$ with 1024-bit preimage and 1024-bit second preimage security up to $2^{64}$-bit messages.

*Better performance:* Keccak-Sp-$\mathrm{F}_{512}$. To achieve the same level of collision and (second) preimage security as SHA-3-512 for messages with at most 64 blocks, we propose Keccak-Sp-$\mathrm{F}_{512}$, an instantiation of $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ with Keccak-p and $h = r' = 512$, $c = 512 + 64 = 576$ and thus $r = 1024$. The throughput of Keccak-Sp-$\mathrm{F}_{512}$ is expected to be 1.77 times that of SHA-3-512. The shortage of Keccak-Sp-$\mathrm{F}_{512}$ is that its indifferentiability is at most $h/2$-bit. However, there *are* numerous important scenarios that only rely on standard (second) preimage security. For example, Schnorr signatures (which remain widely deployed in e.g., Bitcoin BIP340 [90]) rely on properties that are implied by everywhere (second) preimage security of the underlying hashing [72]; RFC 3230 [70] defines a new approach to using message hash digest to achieve a moderate level of integrity; a common practice to achieve file integrity is to upload the hash digest of the file to a trusted third party before distributing the file (see RFC 1805 [81] and 4270 [49, Sect. 3]). The effectiveness of these methods precisely relies on the second preimage resistance of the hash function, to which Keccak-Sp-$\mathrm{F}_{512}$ could be applied.

### 1.3 Discussion

*How to choose $\theta$.* When used with a single fixed IV, the constant $\theta$ can be any non-zero, non-secret value, e.g., one can choose $\theta = 0^{c-1}\|1$. But if two instances of $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ are deployed with different IVs $IV_1, IV_2$ such that $IV_2 = IV_1 \oplus \theta$, a confusion may arise.[8] Namely, a call to $\mathsf{P}(\overline{M}\|IV_1)$ could be both the first permutation-call in $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}(\overline{M}\|\star)$ and the first permutation-call in $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}\big(\mathsf{unpd}(\overline{M})\big)$. This is a general issue in the Merkle-Damgård with permutation construction [48]. Its influence on security is worth investigating. At present, we recommend: (i) When a specification want to use distinct IVs $IV_1, IV_2, ...$ for multiple instances, the IVs and $\theta$ shall be chosen such that $IV_i \neq IV_j \oplus \theta$ for all $i, j$. (ii) If a user wants to use customized IVs, then his IVs shall be randomly chosen to avoid such XOR-relations with other user-customized IVs.

*Quantum security of permutation-based hash functions.* As mentioned, the "invertibility" of the standard sponge construction makes it difficult to establish meaningful quantum security. A series of recent works developed important quantum proof techniques and proved bounds for the standard sponge, but none of them provides meaningful bounds for VIL sponges (or other permutation-based hash functions): (i) [25,63] proved tight preimage bounds for the standard sponge, but for 1 round only; (ii) [1] proved quantum collision and preimage bounds for an arbitrary number of sponge rounds, but concrete security is capped at $\min\{r, c\}/5$ bits; (iii) [28] proved roughly $\min\{c, h\}/2\ell$ bits quantum preimage and $\min\{c, h\}/4\ell$ bits quantum collision bounds for an $\ell$-round sponge. The bound quickly degrades as $\ell$ increases, and becomes meaningless when $\ell \to h$.

More recently, Carolan [24] (concurrently to our work) proved $b/12$-bit preimage and collision bounds for the standard sponge. The $b/12$-bit bottleneck was due to his compressed permutation technique, yielding 133-bit

[8] For example, two variants of Ascon-Hash256 Ascon-XOF128 and Ascon-CXOF128 use $IV_1$ = 0x0000080000cc0003 and $IV_2$ = 0x0000080000cc0004 respectively, which indeed satisfy $IV_1 = IV_2 \oplus \theta$.

security (but no more) for the KECCAK-P parameter $b = 1600$. In all, Carolan's work and this paper appear to give the first permutation-based hash function with "meaningful" quantum security bounds (despite that our proofs are much simpler because of the feed-forward).

*Comparison with Merkle-Damgård using truncated Davies-Meyer.* By adding feed-forward and outputting the inner part of the state, our FOL constructions become quite close to a Merkle-Damgård construction using the truncated Davies-Meyer construction $\mathsf{TrDM}^{\mathsf{P}}_c(X) := \mathsf{right}_c\big(X \oplus \mathsf{P}(X)\big)$ as compression function ($\mathsf{right}_c(X)$ returns the rightmost $c$ bits of $X$). $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ has two advantages. First, the standard Merkle-Damgård construction appends message length to the input to ensure suffix-freeness, which is less efficient than our injective padding for short messages. E.g., in LMS, the majority of hash evaluations have 54 or 55 bytes inputs, on which Merkle-Damgård using truncated Davies-Meyer (and 8-byte encoding of message length, as in SHA-256) would consume one additional permutation-call compared to our construction (and overhead of 14%). Second, by utilizing (instead of discarding) the leftmost $r$ bits of every permutation output, $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ offers better security in some scenarios. Indeed, our proof for lms-mfpspr security of $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ relies on this (Lemma 9 in App. F.2). We also believe keyed variants of $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ yields bounds comparable to the standard keyed sponge, improving over keyed Merkle-Damgård using truncated Davies-Meyer (but formal analysis is left for future work).

*Open questions.* (i) Can we obtain better constructions by adding a feed-forward to [60]; (ii) Is it possible to construct keyless permutations satisfying the CI assumptions in Theorems 2 and 3 from other well-established assumptions; (iii) Is it possible to prove (in some sense) the security of popular permutations (such as the KECCAK-P) w.r.t. the CI assumptions in Theorems 2 and 3.

### 1.4 Related Work

**On Earlier Versions.** Our earlier versions proposed $\mathsf{Sponge\text{-}P}^{\mathsf{P},\mathsf{pd}}$, another variant of sponge-with-feed-forward constructions, which enjoys "IV-freeness", HAIFI-style counters and parallel squeezing. An illustration of its latest version is given in Fig. 12 in App. K. $\mathsf{Sponge\text{-}P}^{\mathsf{P},\mathsf{pd}}$ was developed for the Chinese call for new hash with 1024-bit digests and 1024-bit second preimage security [73]. For lightweight cases (e.g., using ASCON-P), $\mathsf{Sponge\text{-}P}^{\mathsf{P},\mathsf{pd}}$ is inferior to $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$: the counter field is not well compatible with small permutations, and our approach could not prove $c/2$-bit quantum second preimage security. Hence we decided to focus on the IV-based construction $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ in this submission.

**Sponge with Feed-Forward.** Sponge-with-feed-forward is not a fully new idea. In [17], Bhattacharjee et al. proposed to feed the capacity part forward to increase second preimage security w.r.t. random messages. More recently, Sun et al. [88] (which is concurrent and independent to ours) also noticed sponge with feed-forward and proved its advantages, and explored applications to a hash-based signature Ascon-Sign. Differences between our constructions include:

(i) In the absorbing phase, Sun et al.'s construction $\mathsf{SPONGE\text{-}DM}^{\mathsf{P},\mathsf{pd}}$ feeds the *entire b-bit state forward*, whereas we only feed the inner part forward.
(ii) In the squeezing phase, Sun et al.'s $\mathsf{SPONGE\text{-}DM}^{\mathsf{P},\mathsf{pd}}$ does *not* feed forward. In addition, $\mathsf{SPONGE\text{-}DM}^{\mathsf{P},\mathsf{pd}}$ squeezes from the outer part (which follows the standard sponge), whereas we squeeze from the inner part.
(iii) Sun et al. also proposed an Encrypted Davies-Meyer [66] construction-based construction $\mathsf{SPONGE\text{-}EDM}^{\mathsf{P},\mathsf{pd}}$ as their primary proposal.

Due to squeezing from outer part, $\mathsf{SPONGE\text{-}DM}^{\mathsf{P},\mathsf{pd}}$ becomes *less secure* [88, App. B] if it *only feeds the inner part* forward: given a digest, one can evaluate $\mathsf{P}^{-1}$ "backward" for a meet-in-the-middle preimage attack. By squeezing from the already fed-forward inner part, our construction $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ does not have this issue, thus saving $r$ fed-forward bits. Meanwhile, we could have full-inner-squeezing, which: (i) simplifies the CI assumptions; (ii) enables proving meaningful quantum bounds; (iii) increases efficiency for short messages when $h \gg r$ (which is typical in lightweight settings).

For comparison, in App. L we apply our ideas to $\mathsf{SPONGE\text{-}DM}^{\mathsf{P},\mathsf{pd}}$ to reduce its security to CI properties of the permutation. App. L.1 considers case $h = r$ and yields results similar to ours in Sect. 4. App. L.2 briefly discusses $h \gg r$, in which case CI w.r.t. $h/r$-ary relations have to be employed (which resembles the standard sponge) and our method cannot yield satisfactory quantum bounds.

In the context of permutation-based VIL PRFs, Lefevre et al. [58] showed that sponge-with-feed-forward PRF construction allows full state squeezing and achieves forward security, and Eliasi proposed an instance KOALA [2]. This paper proves a similar result: sponge-with-feed-forward hash construction also allows full $c$-bit state squeezing.

**Human Ignorance.** The human ignorance approach was introduced by Rogaway [79], who used it to give a collision security reduction proof for the strengthened Merkle-Damgård construction. Stinson proposed the same idea in his eprint paper [86], and used it to give both collision security reduction and $Z$-preimage security reduction for strengthened Merkle-Damgård. Recently, human ignorance was employed by Mennink [65] to prove security for keyed Merkle-Damgård. Our first step of reducing security of $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ to security of truncated Davies-Meyer $\mathsf{TrDM}_h^{\mathsf{P}}$ resembles [79,86]. However, by using a simpler injective padding *without Merkle-Damgård strengthening-like mechanism*, our collision security reduction has to use $IV$-preimage resistance of $\mathsf{TrDM}_h^{\mathsf{P}}$ as well. By using such a slightly stronger (but still well-acceptable) assumption, we obtain a more efficient construction using injective padding. On the other hand, our reduction proof for second preimage security and the derived bounds in the quantum random permutation model, as well as the reductions to CI of the permutation, are new, compared with [79,86].

Our second proof step appears the first to employ human ignorance for CI. We hope that this could motivate more analogues. For example, in [30, page 216], it was stated that a crucial step to constructing some seedless PRNGs is to develop a keyless UCE notion, on which our approach may shed some lights.

### 1.5 Organization

Sect. 2 notations; Sect. 3 security notions of truncated Davies-Meyer and their relations with CI; Sect. 4 our construction $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ and its security analyses; Sect. 5 a detailed discussion on our findings on LMS; Sect. 6 indifferentiability of $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ *as a XOF*.

Due to page limits, (i) Some proofs for collision, (second) preimage, "LMS-specific" multi-target (second) preimage security and indifferentiability of $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ are deferred to App. E—G; (ii) Our hardware evaluation results are deferred to App. H; (iii) Security proof for LMS using Ascon-Sp-F is deferred to App. I. We also identified a flaw in Fluhrer's [39] in App. I.

## 2 Preliminaries

Let $\{0,1\}^*$ be the set of all finite bit strings, including empty string empty_string. For $X \in \{0,1\}^*$, let $|X|$ denote its bit length. Here $|\text{empty\_string}| = 0$. If $X$ is uniformly distributed over a set $\mathcal{S}$, we write $X \xleftarrow{\$} \mathcal{S}$. For two bit strings $X$ and $Y$, $X\|Y$ is their concatenation. We also write this as $XY$ if it is clear from the context. Let $0^i$ be the string of $i$ zero bits. We write $\mathsf{left}_i(X)$ (resp. $\mathsf{right}_i(X)$) to denote the $i$ leftmost (resp. rightmost) bits of $X$. Given a bit string $X$ of $|X| = \ell r$ bits, let $(X[1], \ldots, X[\ell]) \xleftarrow{r} X$ be the parsing of $X$ into $r$-bit blocks. Here $X[1]\|X[2]\|\ldots\|X[\ell] = X$, $|X[1]| = ...|X[\ell]| = r$.

**Random permutation.** We denote by $\mathcal{P}(b)$ the set of all permutations with domain and range $\{0,1\}^b$. A permutation $\mathbf{\Pi} \xleftarrow{\$} \mathcal{P}(b)$ sampled from $\mathcal{P}(b)$ is a $b$-bit random permutation. Our treatments will use both a standard non-ideal permutation $\mathsf{P}$ and a random permutation $\mathbf{\Pi}$, and we use different notations to make a clear distinction.

**Padding.** Our construction uses an injective padding $\mathsf{pd} : \{0,1\}^* \mapsto (\{0,1\}^r)^*$, which resembles the standard sponge. A special subclass that will be used in Sect. I.2 is *injective appending-padding*, which is defined by appending to the message $M$ a function $sf(M)$ of $M$, such that $\mathsf{apd}(M) := M\|sf(M)$ is injective. An instance of $\mathsf{apd}$ is the classical padding $\mathsf{pd10}^*(M) = M\|10^*$ of [45] that appends a single 1 and then $r - (|M| + 1 \bmod r)$ 0s to $M$. We denote the corresponding unpadding functions by $\mathsf{unpd}$, resp. $\mathsf{unapd}$, and $\mathsf{unpd}(\overline{M})$, resp. $\mathsf{unapd}(\overline{M})$, returns $\perp$ whenever $\overline{M}$ does not correspond to a valid padding.

**Samplable distributions.** Following [11], a distribution $\mathbf{d}$ over a finite set $\mathcal{S}$ is $(t,s)$-*samplable by uniform algorithms*, if there exists a uniform algorithm $\mathscr{S}$ that outputs the string $x$ with probability $\mathbf{d}(x) - \mathbf{d}(x-1)$, where $x-1$ is the lexicographic predecessor of $x$; moreover, $\mathscr{S}$ consumes $t$ time and $s$ space.

**Adversarial resource.** Hash collision and (second) preimage adversaries would output one or two messages. In reduction proofs, the length of the output message(s) typically affects the reduction complexity and thus concrete security, and should be reflected in adversarial resources. We thus call an oracle-aided adversary $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ $(q,t,s,\ell_{\max})$-*bounded*, if: (i) it makes $q$ queries to $\mathsf{P}$ and $\mathsf{P}^{-1}$ in total, consumes $t$ computations and $s$ memory, and (ii) after being padded, every message $M$ output by $\mathscr{A}$ has at most $\ell_{\max}$ blocks. In the plain model (i.e., oracle-free setting), we use $(t,s,\ell_{\max})$-bounded instead. Clearly, $\ell_{\max} = O(t)$, but there may be more stringent restrictions on $\ell_{\max}$ in practice: for example, SHA-256 and SHA-512 have $\ell_{\max} \leq 2^{55}+1$ and $\ell_{\max} \leq 2^{118}+1$ respectively [78].

Subsequently (e.g., in Sect. 2.2), we also consider adversaries that do not yield long outputs, and we use $(q,t,s)$-bounded and $(t,s)$-bounded instead.

### 2.1 Collision and (Second) Preimage Security of Hashing

This section presents the traditional security definitions for permutation oracle-based fixed-output-length hash functions $\mathsf{h}^{\mathsf{P}}$.

First, given an adversary $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$, we define the attack advantage of $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ against the collision security of $\mathsf{h}^{\mathsf{P}}$ as follows:

$$\mathbf{Adv}^{\mathrm{coll}}_{\mathsf{h}}(\mathscr{A}) := \mathrm{Pr}_{\mathsf{P},\varepsilon}\big[\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}} \Rightarrow (M,M') : M \neq M',\ \mathsf{h}^{\mathsf{P}}(M) = \mathsf{h}^{\mathsf{P}}(M')\big], \tag{1}$$

where the probability is taken over the choice of $\mathsf{P}$ (if $\mathsf{P}$ has randomness) and the randomness $\varepsilon$ of $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$.

Finding a preimage for $\mathsf{h}^{\mathsf{P}}$ requires finding $M \in \{0,1\}^*$ for a challenge image $Z \in \{0,1\}^h$, such that $\mathsf{h}^{\mathsf{P}}(M) = Z$. Given $\mathsf{h}^{\mathsf{P}}$ and $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ and a challenge image $Z \in \{0,1\}^h$, we define the advantage of $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ against the $Z$-preimage security of $\mathsf{h}^{\mathsf{P}}$ as follows:

$$\mathbf{Adv}^{Z\text{-pre}}_{\mathsf{h}}(\mathscr{A}) := \mathrm{Pr}_{\mathsf{P},\varepsilon}\big[\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}(Z) \Rightarrow M \in \{0,1\}^*,\ \mathsf{h}^{\mathsf{P}}(M) = Z\big]. \tag{2}$$

Then, the advantage of $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ against the *everywhere preimage* security of $\mathsf{h}^{\mathsf{P}}$ is defined by taking maximum over all image $Z$.

$$\mathbf{Adv}^{\mathrm{epre}}_{\mathsf{h}}(\mathscr{A}) := \max_{Z\in\{0,1\}^h}\Big\{\mathbf{Adv}^{Z\text{-pre}}_{\mathsf{h}}(\mathscr{A})\Big\}. \tag{3}$$

Finding a second preimage for $\mathsf{h}^{\mathsf{P}}$ for a challenge message $M \in \{0,1\}^*$ requires finding $M' \in \{0,1\}^*$ such that $M' \neq M$ while $\mathsf{h}^{\mathsf{P}}(M') = \mathsf{h}^{\mathsf{P}}(M)$. Concretely, given $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ and a challenge message $M \in \{0,1\}^*$, we define the advantage of $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ against the $M$-second preimage security of $\mathsf{h}^{\mathsf{P}}$ as follows:

$$\begin{aligned}\mathbf{Adv}^{M\text{-spr}}_{\mathsf{h}}(\mathscr{A}) := \mathrm{Pr}_{\mathsf{P},\varepsilon}\big[\ &\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}(M) \Rightarrow M' \in \{0,1\}^*,\ M' \neq M,\\ &\mathsf{h}^{\mathsf{P}}(M') = \mathsf{h}^{\mathsf{P}}(M)\big].\end{aligned} \tag{4}$$

Typically, second preimage resistance is then defined w.r.t. the bit-length of the challenge message $M$ [80,91], which influences concrete security (as observed in [54,88]). Formally, the advantage of $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ against the everywhere second preimage security of $\mathsf{h}^{\mathsf{P}}$ for $bl$-bit challenge messages is:

$$\mathbf{Adv}^{\mathrm{espr}[bl]}_{\mathsf{h}}(\mathscr{A}) := \max_{M\in\{0,1\}^{bl}}\Big\{\mathbf{Adv}^{M\text{-spr}}_{\mathsf{h}}(\mathscr{A})\Big\}. \tag{5}$$

**Impossibility of Everywhere-style Notions for Keyless Hash.** Note that *keyless, non-idealized hash functions never achieve everywhere preimage security.* This was also informally mentioned in [7]. The reason is as follows. Consider a keyless $\mathsf{h}$ that is not defined upon any ideal primitive. Recall from Eq. (3) that $\mathbf{Adv}^{\mathrm{epre}}_{\mathsf{h}}(\mathscr{A}) = \max_{Z\in\{0,1\}^h}\big\{\mathbf{Adv}^{Z\text{-pre}}_{\mathsf{h}}(\mathscr{A})\big\}$. By this, $\mathscr{A}$ simply outputs an arbitrary fixed message $M^\circ \in \{0,1\}^*$, and this will cause $\mathbf{Adv}^{\mathsf{h}(M^\circ)\text{-pre}}_{\mathsf{h}}(\mathscr{A}) = 1$. By this, the maximum over $Z \in \{0,1\}^h$ returns probability 1 as well. The same impossibility holds for everywhere second preimage resistance $\mathbf{Adv}^{\mathrm{espr}[bl]}_{\mathsf{h}}(\mathscr{A})$ (as long as the output message $M^\circ \in \{0,1\}^{bl}$ has a second preimage in $\{0,1\}^{bl}$). Unlike the well-known non-uniform collision adversary against keyless hash functions, this attack is *uniform* and applicable to *all* hash functions with domain $Dom \ni M^\circ$.

By these, everywhere-style notions cannot be security goals nor assumptions for non-idealized, keyless hash functions. As elaborated below, there are two approaches to weaken these notions.

**Distribution-oriented (Second) Preimage Resistance.** For keyless hash functions, Rogaway and Shrimpton [80] introduced *always preimage resistance*, which requires $\mathscr{A}(Z)$ to find the preimage for $Z = \mathsf{h}(M)$ for a uniformly picked message $M \xleftarrow{\$} \{0,1\}^{bl}$. This is called *domain-oriented preimage resistance* in [7]. As its dual, *range-oriented preimage resistance* [7] requires $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}(Z)$ to find the preimage for a uniformly picked image $Z \xleftarrow{\$} \{0,1\}^h$. Andreeva and Stam [7] further extended these notions to *distribution-oriented*, by requiring the challenge image $Z$ to be sampled according to a distribution $\mathbf{d}$ (which then becomes a parameter of the security definition). In the keyless setting, by allowing the distribution to depend on the function $\mathsf{h}$, domain-distribution-oriented and range-distribution-oriented variants can be unified: a distribution over $\{0,1\}^*$ can be transformed into an $\mathsf{h}$-dependent distribution over $\{0,1\}^h$.

Concretely, (in the plain model,) given a distribution $\mathbf{d}$ over $\{0,1\}^h$, the advantage of $\mathscr{A}$ against the $\mathbf{d}$-oriented preimage security of $\mathsf{h}$ is

$$\begin{aligned}\mathbf{Adv}^{\mathbf{d}\text{-pre}}_{\mathsf{h}}(\mathscr{A}) &:= \Pr_{\varepsilon}\big[Z \xleftarrow{\$} \mathbf{d},\ \mathscr{A}(Z) \Rightarrow M \in \{0,1\}^*,\ \mathsf{h}^{\mathsf{P}}(M) = Z\big] \\ &= \sum_{Z\in\{0,1\}^h} \Pr\big[Z^* \xleftarrow{\$} \mathbf{d} : Z^* = Z\big] \cdot \mathbf{Adv}^{Z\text{-pre}}_{\mathsf{h}}(\mathscr{A}). \end{aligned} \tag{6}$$

As will be seen, regarding preimage resistance, distribution-oriented notion will be an interpretation of our main theorem (Theorem 2); regarding second preimage, we only use a distribution-oriented form as the final result (Theorem 3).

Similarly, one can define *distribution-oriented second preimage resistance*: given a distribution $\mathbf{d}$ over $\{0,1\}^*$, the advantage of $\mathscr{A}$ against the $\mathbf{d}$-oriented second preimage security of $\mathsf{h}$ is

$$\mathbf{Adv}^{\mathbf{d}\text{-spr}}_{\mathsf{h}}(\mathscr{A}) := \sum_{M\in\{0,1\}^*} \Pr\big[M^* \xleftarrow{\$} \mathbf{d} : M^* = M\big] \cdot \mathbf{Adv}^{M\text{-spr}}_{\mathsf{h}}(\mathscr{A}). \tag{7}$$

### 2.2 Correlation Intractability for Permutations

An $m$-ary relation $R$ is a subset $\mathcal{S} \subseteq (\{0,1\}^b)^m \times (\{0,1\}^b)^m$. Given a keyless permutation $\mathsf{P} : \{0,1\}^b \mapsto \{0,1\}^b$, a relation $R$ and an adversary $\mathscr{A}$, the *$R$-correlation intractability ($R$-ci) advantage of $\mathscr{A}$ against $\mathsf{P}$ w.r.t. $R$* is

$$\begin{aligned}\mathbf{Adv}^{R\text{-ci}}_{\mathsf{P}}(\mathscr{A}) := \Pr_{\mathsf{P},\varepsilon}\big[\ &\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}(R) \Rightarrow (X_1,...,X_m), \\ &\big((X_1,...,X_m),(\mathsf{P}(X_1),...,\mathsf{P}(X_m))\big) \in R\big],\end{aligned} \tag{8}$$

where the probability is taken over the randomness $\varepsilon$ of $\mathscr{A}$.

Not all relations are "meaningful". E.g., if $(x,y) \in R$ for all $x,y \in \{0,1\}^b$, then no permutation $\mathsf{P}$ can have a small advantage $\mathbf{Adv}^{R\text{-ci}}_{\mathsf{P}}(\mathscr{A})$ for this "trivial" relation $R$. To quantify the "non-triviality", a relation $R$ is *evasive* (in the asymptotic sense), if $\mathbf{Adv}^{R\text{-ci}}_{\mathbf{\Pi}}(\mathscr{A})$ is negligible for every efficient adversary $\mathscr{A}$. As will be seen in Sect. 4.3, the relations for measuring the "goodness" of $\mathsf{P}$ in this paper are all evasive.

Being $R$-ci for *every* evasive relation $R$ was introduced in [23] as a formalism for "being random-oracle-like". To fit into the non-uniform adversary model, existing works considered *keyed permutations* [23] (requiring a permutation picked from a family to be CI against non-uniform adversaries) or idealized permutations [27,83]. Even for keyed permutations, being $R$-ci for any evasive $R$ is not achievable in the standard model [23], but exciting positive results have been achieved regarding restricted relations [21,52,22,20,76,62,26] (e.g., relations recognizable with bounded complexity [21]). We refer to App. B for more details. Our reduction proofs will rely on very specific and simple relations $R$ (see Sect. 3), so that $R$-ci seems achievable and testable by cryptanalytic practice.

### 2.3 Quantum

Our quantum bounds are derived from the classical proofs and a recent quantum lifting theorem of [28]. Readers thus do not need detailed background on quantum computing, and here we only present necessary features (for completeness, a brief background is given in App. C).

A quantum adversary is an algorithm (uniform model) or a circuit (non-uniform) operating on *qubits*, which are the basic storage element in the quantum world. The state of $b$ qubits includes *all* $X \in \{0,1\}^b$ in *superposition* (by quantum mechanism), so that an operation over them can compute over all $X \in \{0,1\}^b$ simultaneously. Similarly, a quantum oracle for a classical permutation $\mathsf{P}$ has queries and responses in superpositions as well, and a single query/response could yield $\mathsf{P}(X)$ for *all* $X \in \{0,1\}^b$. However, measuring a superposition only probabilistically yields one of the $2^b$ values. To have the desired result (e.g., a collision $\mathsf{h}(M) = \mathsf{h}(M')$) w.h.p., dedicated techniques have been deployed [43]: for any hash function with $h$-bit outputs, quantum algorithms can find a collision within roughly $2^{h/3}$ steps and find the preimage for a certain $Z \in \{0,1\}^h$ within $2^{h/2}$ steps, both of which are faster than the classical algorithms. As a result, it is desirable for a $h$-bit hash construction to achieve $h/3$ bit quantum collision and $h/2$-bit quantum preimage provable security.

Each of our attack advantage definitions (including Eqs. (1), (2), (3), (4), (5), (8), etc.) can be extended to a corresponding quantum analogue, by considering quantum adversaries issuing superposition queries to

its oracles. We add a prefix "q-" to highlight: e.g., we use $\mathbf{Adv}^{\text{q-coll}}_{\text{XOF}}(\mathscr{A})$ to denote quantum collision attack advantage of a quantum adversary $\mathscr{A}$ against XOF.

A recent work of Cojocaru et al. [28] proved theorem lifting evasiveness w.r.t. classical random permutations to that w.r.t. quantum random permutations.

**Theorem 1 ([28], Theorem 4.1).** *Let $\mathscr{A}^{\boldsymbol{\Pi},\boldsymbol{\Pi}^{-1}}$ be a quantum algorithm that makes $q$ quantum queries to an (invertible) random permutation oracle $\boldsymbol{\Pi} \xleftarrow{\$} \mathcal{P}(b)$ and $R$ is a relation on $(\{0,1\}^b)^k \times (\{0,1\}^b)^k$. Then there exists an algorithm $\mathscr{B}^{\boldsymbol{\Pi},\boldsymbol{\Pi}^{-1}}$ making at most $k$ classical queries such that*

$$\begin{aligned}&\Pr_{\boldsymbol{\Pi}}\big[\mathscr{B}^{\boldsymbol{\Pi},\boldsymbol{\Pi}^{-1}} \Rightarrow (X_1,...,X_k) : \big((X_1,...,X_k),(\boldsymbol{\Pi}(X_1),...,\boldsymbol{\Pi}(X_k))\big) \in R\big]\\ &\geq \frac{\left(1-\frac{k^2}{2^b}\right)}{(8q+1)^{2k}} \cdot \Pr_{\boldsymbol{\Pi}}\big[\mathscr{A}^{\boldsymbol{\Pi},\boldsymbol{\Pi}^{-1}} \Rightarrow (X_1,...,X_k) :\\ &\qquad\qquad \big((X_1,...,X_k),(\boldsymbol{\Pi}(X_1),...,\boldsymbol{\Pi}(X_k))\big) \in R\big]. \end{aligned} \tag{9}$$

## 3 Security of Truncated Davies-Meyer, and CI of P

Built upon the permutation $\mathsf{P}$, the truncated Davies-Meyer construction $\mathsf{TrDM}^{\mathsf{P}}_{\nu}$ is defined as $\mathsf{TrDM}^{\mathsf{P}}_{\nu}(X) := \mathsf{right}_{\nu}\big(X \oplus \mathsf{P}(X)\big)$. We will first reduce security of $\mathsf{Sponge}\text{-}\mathsf{F}^{\mathsf{P},\mathsf{pd}}$ to security of $\mathsf{TrDM}^{\mathsf{P}}_{h}$, which then is reduced to $R$-ci of $\mathsf{P}$ for some relation $R$ picked from very restricted sets.

**Security of $\mathsf{TrDM}^{\mathsf{P}}_{\nu}$, and relations $R_{\text{coll}}$, $R_{\text{pre}}(Z)$ and $R_{\text{spr}}(X,\nu)$.** As a folklore, finding a collision on $\mathsf{TrDM}^{\mathsf{P}}_{h}$ indicates finding two inputs $X, X'$ of $\mathsf{P}$ satisfying the following binary relation:

$$\begin{aligned}&\big((X,X'),(\mathsf{P}(X),\mathsf{P}(X'))\big) \in R_{\text{coll}}\\ &\qquad \textbf{iff } X \neq X' \wedge \big(\mathsf{right}_h(X \oplus \mathsf{P}(X)) = \mathsf{right}_h(X' \oplus \mathsf{P}(X'))\big).\end{aligned} \tag{10}$$

On the other hand, finding a preimage $(\mathsf{TrDM}^{\mathsf{P}}_{\nu})^{-1}(Z)$ for $Z \in \{0,1\}^{\nu}$ is equivalent with finding $X$ satisfying the following relation:

$$(X, \mathsf{P}(X)) \in R_{\text{pre}}(Z) \textbf{ iff } \mathsf{right}_{|Z|}(X \oplus \mathsf{P}(X)) = Z. \tag{11}$$

Our analysis only needs two particular choices of $\nu$, i.e., $\nu = c$ and $\nu = h$. With these in mind, we further introduce a family of relations indexed by $X$ and $\nu$:

$$\mathcal{R}_{\text{pre}} := \big\{R_{\text{pre}}(Z),\ X \in (\{0,1\}^c \cup \{0,1\}^h)\big\}. \tag{12}$$

If $\mathsf{P}$ is $R_{\text{pre}}(Z)$-ci for all $R_{\text{pre}}(Z) \in \mathcal{R}_{\text{pre}}$, then $\mathsf{TrDM}^{\mathsf{P}}_{h}$ might be everywhere preimage resistant. Unfortunately, when $\mathsf{P}$ is keyless and non-idealize, this "everywhere-style" ci property is not achievable either.

Finally, given a challenge input $X \in \{0,1\}^b$ of $\mathsf{TrDM}^{\mathsf{P}}_{\nu}$, an adversary finding a second preimage $\mathsf{TrDM}^{\mathsf{P}}_{\nu}(X') = \mathsf{TrDM}^{\mathsf{P}}_{\nu}(X)$ yields an adversary finding $X'$ satisfying the following relation:

$$(X', \mathsf{P}(X')) \in R_{\text{spr}}(X,\nu) \textbf{ iff } X' \neq X,\ \mathsf{right}_{\nu}(X \oplus \mathsf{P}(X)) = \mathsf{right}_{\nu}(X' \oplus \mathsf{P}(X')). \tag{13}$$

Built on these, we introduce the following relation family indexed by $X$ and $\nu$:

$$\mathcal{R}_{\text{spr}} := \big\{R_{\text{spr}}(X,\nu),\ X \in \{0,1\}^b,\ \nu \in \{c,h\}\big\}. \tag{14}$$

**Multi-target second preimage security and $R_{\text{spr}}(X,\nu)$-ci.** We will reduce the second preimage security of $\mathsf{Sponge}\text{-}\mathsf{F}^{\mathsf{P},\mathsf{pd}}$ to special variants of *multi-target* second preimage security of $\mathsf{TrDM}^{\mathsf{P}}_{c}$ and $R_{\text{spr}}(X,\nu)$-ci (as mentioned in the Introduction), which are defined as follows.

First, given $\ell$ challenge inputs $X_1,...,X_\ell \in \{0,1\}^b$, the advantage of $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ against the $(X_1,...,X_\ell)$-multi-target second preimage security of $\mathsf{TrDM}^{\mathsf{P}}$ is

$$\begin{aligned}&\mathbf{Adv}^{(X_1,...,X_\ell)\text{-mspr}^*}_{\mathsf{TrDM}^{\mathsf{P}}}(\mathscr{A})\\ &:= \Pr_{\mathsf{P},\varepsilon}\big[\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}(X_1,...,X_\ell) \Rightarrow X' :\ \mathsf{TrDM}^{\mathsf{P}}_{h}(X') = \mathsf{TrDM}^{\mathsf{P}}_{h}(X_\ell) \wedge X' \neq X_\ell,\ \text{or}\\ &\qquad \mathsf{TrDM}^{\mathsf{P}}_{c}(X') = \mathsf{TrDM}^{\mathsf{P}}_{c}(X_i) \text{ for some } i \in \{1,...,\ell-1\} \wedge X' \neq X_i\big].\end{aligned} \tag{15}$$

Namely, $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ succeeds as long as it finds: (i) a second preimage for the $h$-bit function $\mathsf{TrDM}_h^{\mathsf{P}}$ for the challenge $X_\ell$, or (ii) a second preimage for the $c$-bit function $\mathsf{TrDM}_c^{\mathsf{P}}$ for any of the $\ell-1$ challenges $X_1, ..., X_{\ell-1}$.

Similarly, given $\ell$ relations $R_1, ..., R_\ell \in \mathcal{R}_{\rm spr}$, the advantage of $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ against the $(R_1, ..., R_\ell)$-multi-target correlation intractability of $\mathsf{P}$ is

$$\mathbf{Adv}_{\mathsf{P}}^{(R_1,...,R_\ell)\text{-mci}}(\mathscr{A}) := \Pr_{\mathsf{P},\varepsilon}\big[\ \mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}(R_1, ..., R_\ell) \Rightarrow X', \\ (X', \mathsf{P}(X')) \in R_i \text{ for some } i \in \{1, ..., \ell\}\big]. \tag{16}$$

In our reduction, the first $\ell-1$ relations $R_1, ..., R_{\ell-1}$ will be picked from $R_{\rm spr}(\star, c)$, whereas the last relation $R_\ell$ will be picked from $R_{\rm spr}(\star, h)$.

**Distribution-oriented CI.** Similarly to the case of everywhere preimage, there exists an uniform adversary $\mathscr{A}$ such that the "everywhere-style" CI advantage $\max_{R\in\mathcal{R}_{\rm pre}}\left\{\mathbf{Adv}_{\mathsf{P}}^{R\text{-ci}}(\mathscr{A})\right\}$ is 1: $\mathscr{A}$ just outputs $0^b$ to make $\mathbf{Adv}_{\mathsf{P}}^{\mathsf{P}(0^b)\text{-ci}}(\mathscr{A}) = 1$.

We thus follow Sect. 2.1 to define *distribution-oriented CI*. Concretely, given a distribution $\mathbf{d}$ over a relation family $\mathcal{R}$, the advantage of $\mathscr{A}$ against the $\mathbf{d}$-oriented $\mathcal{R}$-ci security of $\mathsf{P}$ is

$$\mathbf{Adv}_{\mathsf{P}}^{\mathbf{d}\text{-}\mathcal{R}\text{-ci}}(\mathscr{A}) := \sum_{R\in\mathcal{R}} \Pr\big[R^* \xleftarrow{\$} \mathbf{d} : R^* = R\big] \cdot \mathbf{Adv}_{\mathsf{P}}^{R\text{-ci}}(\mathscr{A}). \tag{17}$$

The challenge tuple $(X_1, ..., X_\ell)$ in Eq. (15) and $(R_1, ..., R_\ell)$ in Eq. (16) may be sampled according to distributions as well, leading to *distribution oriented multi-target second preimage resistance* and *distribution oriented multi-target CI*:

$$\mathbf{Adv}_{\mathsf{P}}^{\mathbf{d}\text{-mspr}^*}(\mathscr{A}) := \sum_{(X_1,...,X_\ell)\in(\{0,1\}^b)^\ell} \Pr\big[\ (X_1^*, ..., X_\ell^*) \xleftarrow{\$} \mathbf{d} : (X_1^*, ..., X_\ell^*) = (X_1, ..., X_\ell)\big] \\ \cdot \mathbf{Adv}_{\mathsf{TrDM}_c^{\mathsf{P}}}^{(X_1,...,X_\ell)\text{-mspr}^*}(\mathscr{A}), \tag{18}$$

$$\mathbf{Adv}_{\mathsf{P}}^{\mathbf{d}\text{-}\mathcal{R}\text{-mci}}(\mathscr{A}) := \sum_{(R_1,...,R_\ell)\in\mathcal{R}^\ell} \Pr\big[\ (R_1^*, ..., R_\ell^*) \xleftarrow{\$} \mathbf{d} : (R_1^*, ..., R_\ell^*) = (R_1, ..., R_\ell)\big] \\ \cdot \mathbf{Adv}_{\mathsf{P}}^{(R_1,...,R_\ell)\text{-}\mathcal{R}\text{-mci}}(\mathscr{A}). \tag{19}$$

These will be the foundation of Theorem 3.

**Equivalence of $\mathsf{TrDM}_h^{\mathsf{P}}$ security and $\mathsf{P}$ CI.** A folklore is that the collision and (second) preimage security of the (truncated) Davies-Meyer construction $\mathsf{TrDM}_h^{\mathsf{P}}$ security are equivalent with CI of $\mathsf{P}$. We present a formal claim as follows.

**Lemma 1.** *Let $(\mathsf{P}, \mathsf{P}^{-1})$ be a permutation oracle (not necessarily random) and its inverse. Then, for $\mathsf{X} \in \{$classical, quantum$\}$, for all $Z \in \{0,1\}^h$ and $X \in \{0,1\}^b$, any $\mathsf{X}$-adversary $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ breaking the $\mathsf{A}$ security of $\mathsf{TrDM}_h^{\mathsf{P}}$ can be used to break the $\mathsf{B}$ security of $\mathsf{P}$ with the same attack advantage, i.e.,*

$$\mathbf{Adv}_{\mathsf{TrDM}_h^{\mathsf{P}}}^{\mathsf{A}}(\mathscr{A}) = \mathbf{Adv}_{\mathsf{P}}^{\mathsf{B}}(\mathscr{A}),$$

*where* $(\mathsf{A}, \mathsf{B}) \in \big\{(Z\text{-pre}, R_{\rm pre}(Z)\text{-ci}),\ (X\text{-spr}, R_{\rm spr}(X,h)\text{-ci}),\ (\text{coll}, R_{\rm coll}\text{-ci})\big\}$.

A short proof is given in Appendix E.1.

## 4 Sponge-$\mathsf{F}^{\mathsf{P},\mathsf{pd}}$ and Its Security

This section presents our first construction Sponge-$\mathsf{F}^{\mathsf{P},\mathsf{pd}}$ and its security analyses. The description is given in Sect. 4.1, which is followed by collision and (second) preimage security reductions in Sect. 4.2 and precise bounds in the (quantum) random permutation model in Sect. 4.3.

### 4.1 Description of Sponge-$\mathsf{F}^{\mathsf{P},\mathsf{pd}}$

Sponge-$\mathsf{F}^{\mathsf{P},\mathsf{pd}}$ is defined upon a permutation $\mathsf{P} : \{0,1\}^b \mapsto \{0,1\}^b$ and an injective padding function $\mathsf{pd} : \{0,1\}^* \mapsto (\{0,1\}^r)^{\ell_{\max}}$. Let $M \in \{0,1\}^*$ and $\nu$ be an integer such that $h \le \nu \le \nu_{\max}$. On input $(M, \nu)$, Sponge-$\mathsf{F}^{\mathsf{P},\mathsf{pd}}(M, \nu)$ outputs a string $Z \in \{0,1\}^\nu$. Besides the minimal output size $h$, Sponge-$\mathsf{F}^{\mathsf{P},\mathsf{pd}}$ has three additional parameters $r, c, r' \in \mathbb{N}$ and a fixed public constant $IV \in \{0,1\}^c$, where $c \ge r' \ge h$, $r + c = b$, where $r$, resp. $c$, is the bit length of the "sponge" outer, resp. inner, part of the $b$-bit internal state. The concrete process of Sponge-$\mathsf{F}^{\mathsf{P},\mathsf{pd}}$ is described by the algorithms in Fig. 3 and illustrated in Fig. 2.

**Procedure** $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}(M,\nu)$
$(\overline{M}[1],\ldots,\overline{M}[\ell]) \xleftarrow{r} \mathsf{pd}(M)$, $X \leftarrow 0^r\|IV$
// Absorbing phase
**for** $i = 1,2,...,\ell$ **do**
  **if** $i \neq \ell$ **then**
    $X \leftarrow X \oplus (\overline{M}[i]\|0^c)$
  **else** // $i = \ell$
    $X \leftarrow X \oplus (\overline{M}[i]\|\theta)$
  $S \leftarrow \mathsf{right}_c(X)$, $X \leftarrow \mathsf{P}(X) \oplus (0^r\|S)$
// Squeezing phase
$\lambda \leftarrow \lceil \frac{\nu}{r'} \rceil$
**for** $i = 1,2,...,\lambda$ **do**
  $Z[i] \leftarrow \mathsf{right}_{r'}(X)$
  $S \leftarrow \mathsf{right}_{r'}(X)$, $X \leftarrow \mathsf{P}(X) \oplus (0^r\|S)$
**return** $\mathsf{left}_\nu\big(Z[1]\|...\|Z[\lambda]\big)$

**Fig. 3.** $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$: the XOF construction built upon the sponge with feed-forward construction, with parameters $r, c, r'$ and injective padding $\mathsf{pd} : \{0,1\}^* \mapsto (\{0,1\}^r)^*$.

**Reducing fed-forward bits?** Intuitively, reducing feed-forward bits (and thus costs) to be less than capacity for *all* permutation-calls would reduce concrete security. The extreme case is the standard sponge, in which the number of feed-forward bits is zero. Still, it is an interesting future work to characterize the influence of the number of feed-forward bits, when less than the capacity, on concrete security.

A natural question is whether we can reduce feed-forward bits for some (but not all) permutation-calls to "steal" some efficiency *without sacrificing security*. Our opinions are two-fold:

- First, in common situations where all permutation-calls are executed on the same device, this does *not* improve efficiency, because one has to reserve registers for the largest fed-forward. Concretely, let's informally assume that the $i$-th permutation-call has $c_i$ input bits fed-forward. Then, the implementation would always have to reserve $\max_i\{c_i\}$ bits registers, and reducing feed-forward bits of the other calls would *not* reduce this cost. Actually, this memory constitutes the main cost of feed-forward. Our paper thereby focuses on *always* feeding the $c$-bit inner part of the inputs forward.
- On the other hand, there might be cases where absorption can use a larger footprint, and vice versa. E.g., a user may delegate the absorption to a server and obtain the $b$-bit final state, and then squeeze using its constrained device.
  This application can use a variant of XOF version of $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ (Fig. 2 Top), using $c$-bit capacity and feed-forward in absorption and $r' < c$ bit capacity, feed-forward and squeezing rate in squeezing. It needs an $h/r'$-ary relation for (second) preimage security reduction and an $2h/r'$-ary relation for collision security. In classical setting, the same bounds as our current results can still be proven; but Theorem 1 yields much weaker quantum bounds due to the increased relation ary.
  These use cases as well as $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ variants for them are also interesting future works.

### 4.2 (Second) Preimage and Collision Resistance of $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$

In this subsection, we *fix the second input of* $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ *to* $h$ and view it as a conventional fixed-output-length hash $\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}(\cdot) := \mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}(\cdot,h)$, and analyze it w.r.t. the security definitions in Sect. 2.1.

**Reducing to security of $\mathsf{TrDM}_h^{\mathsf{P}}$ and CI of $\mathsf{P}$.** We first show that breaking the collision security of $\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}$ means breaking either the collision security of $\mathsf{TrDM}_h^{\mathsf{P}}$ or the preimage security of $\mathsf{TrDM}_c^{\mathsf{P}}$ for the image $IV$. This resembles Lemma 7.1 of [10]. We consider a more general setting where $(\mathsf{P},\mathsf{P}^{-1})$ are provided as permutation oracles, which would enable deriving results in both the plain model and the (quantum) random permutation model.

**Lemma 2 (Collision).** *Let* $(\mathsf{P},\mathsf{P}^{-1})$ *be a permutation oracle (not necessarily random) and its inverse,* $\mathsf{X} \in \{$*classical*, *quantum*$\}$*, and let* $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ *be a* $(q,t,s,\ell_{\max})$*-bounded* $\mathsf{X}$ *collision adversary against* $\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}$*.*

*Then, we can construct a* $(q+2\ell_{\max}, t+O(\ell_{\max}), s+O(\ell_{\max}))$-$\mathsf{X}$ *collision adversary* $\mathscr{B}_{\mathrm{coll}}^{\mathsf{P},\mathsf{P}^{-1}}$ *against* $\mathsf{TrDM}_h^{\mathsf{P}}$ *and a* $(q+2\ell_{\max}, t+O(\ell_{\max}), s+O(\ell_{\max}))$*-preimage adversary* $\mathscr{B}_{\mathrm{pre}}^{\mathsf{P},\mathsf{P}^{-1}}$ *against* $\mathsf{TrDM}_c^{\mathsf{P}}$*, such that:*

$$\mathbf{Adv}^{\mathrm{coll}}_{\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}}(\mathscr{A}) \le \mathbf{Adv}^{\mathrm{coll}}_{\mathsf{TrDM}_h^{\mathsf{P}}}(\mathscr{B}_{\mathrm{coll}}) + \mathbf{Adv}^{IV\text{-}\mathrm{pre}}_{\mathsf{TrDM}_c^{\mathsf{P}}}(\mathscr{B}_{\mathrm{pre}}) \tag{20}$$

$$= \mathbf{Adv}^{R_{\mathrm{coll}}\text{-}\mathrm{ci}}_{\mathsf{P}}(\mathscr{B}_{\mathrm{coll}}) + \mathbf{Adv}^{R_{\mathrm{pre}}(IV)\text{-}\mathrm{ci}}_{\mathsf{P}}(\mathscr{B}_{\mathrm{pre}}). \tag{21}$$

The proof is given in Appendix E.2.

Similarly, finding a preimage of $\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}$ for $Z$ means finding a preimage of $\mathsf{TrDM}_h^{\mathsf{P}}$ for $Z$. We refer to Appendix E.3 for a formal elaboration.

We finally consider second preimage, which is a bit complicated. We define two properties for a second preimage challenge $M \in \{0,1\}^*$:

- $M$ is *free-of-inner-collision*, if the permutation calls $\mathsf{P}(X_1) = Y_1, ..., \mathsf{P}(X_\ell) = Y_\ell$ underlying $\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}(M)$ are such that $\mathsf{right}_h(X_1 \oplus Y_1), ..., \mathsf{right}_h(X_\ell \oplus Y_\ell)$ are distinct;
- $M$ is *free-of-IV-preimage*, if the calls $\mathsf{P}(X_1) = Y_1, ..., \mathsf{P}(X_\ell) = Y_\ell$ underlying $\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}(M)$ are such that $IV \notin \{\mathsf{right}_c(X_1 \oplus Y_1), ..., \mathsf{right}_c(X_\ell \oplus Y_\ell)\}$.

Roughly, being free-of-inner-collision means $M$ does not yield a trivial second preimage for itself. E.g., if $\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}(M)$, $\mathsf{pd}(M) = \overline{M}[1]\|\overline{M}[2]\|\overline{M}[3]\|\overline{M}[4]$, has calls $\mathsf{P}(X_1) = Y_1, ..., \mathsf{P}(X_4) = Y_4$ with $\mathsf{right}_c(X_1 \oplus Y_1) = \mathsf{right}_c(X_3 \oplus Y_3)$, then it can be seen $\mathsf{unpd}\big(\overline{M}[1]\|\overline{M}[2]\|\overline{M}[3]\|\mathsf{left}_r(Y_3) \oplus \mathsf{left}_r(X_2)\|\overline{M}[3]\|\overline{M}[4]\big)$, if not being $\perp$, is a second preimage of $M$. Such a challenge $M$ is thus meaningless. We remark that this example concerns with $\mathsf{right}_c(X_i \oplus Y_i), i = 1, ..., \ell$, but our definition of free-of-inner-collision requires distinctness of the shorter strings $\mathsf{right}_h(X_i \oplus Y_i), i = 1, ..., \ell$ for simplicity. Whereas, being free-of-IV-preimage means $M$ cannot help compute $(\mathsf{TrDM}_c^{\mathsf{P}})^{-1}(IV)$.

We then show that if $M$ is both free-of-inner-collision and free-of-IV-preimage, then finding a second preimage of $\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}$ for $M$ basically means breaking multi-target second preimage security of $\mathsf{TrDM}_c^{\mathsf{P}}$ for a corresponding tuple of $b$-bit challenge inputs.

**Lemma 3 (Second Preimage).** *Let* $(\mathsf{P}, \mathsf{P}^{-1})$ *be a permutation oracle (not necessarily random) and its inverse. We can construct a classical procedure* $\textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}$ *such that: for every challenge message* $M \in \{0,1\}^{bl}$ *that is both free-of-inner-collision and free-of-IV-preimage, and every* $(q, t, s, \ell_{\max})$*-bounded* $\mathsf{X}$ *second preimage adversary* $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ *against* $\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}$ *for the challenge* $M$*,* $\mathsf{X} \in \{classical, quantum\}$*, the procedure* $\textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}(M) \Rightarrow (X_1, ..., X_\ell)$ *outputs* $\ell \le \ell_{\max}$ *distinct challenge inputs in* $\{0,1\}^b$*, and we can construct:*

*(i)* *A* $(q+2\ell_{\max}, t_A+O(\ell_{\max}), s_A+O(\ell_{\max}))$-$\mathsf{X}$ *multi-target second preimage adversary* $\mathscr{B}_{\mathrm{mspr}^*}^{\mathsf{P},\mathsf{P}^{-1}}$ *against* $\mathsf{TrDM}^{\mathsf{P}}$ *for the challenge inputs* $(X_1, ..., X_{\ell-1})$*;*

*(ii)* *A* $(q+2\ell_{\max}, t_A+O(\ell_{\max}), s_A+O(\ell_{\max}))$*-preimage adversary* $\mathscr{B}_{\mathrm{pre}}^{\mathsf{P},\mathsf{P}^{-1}}$ *against* $\mathsf{TrDM}_c^{\mathsf{P}}$ *for the challenge image* $IV$*.*

*And for* $(R_1, ..., R_{\ell-1}, R_\ell) = (R_{\mathrm{spr}}(X_1, c), ..., R_{\mathrm{spr}}(X_{\ell-1}, c), R_{\mathrm{spr}}(X_\ell, h))$*, it holds:*

$$\mathbf{Adv}^{M\text{-}\mathrm{spr}}_{\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}}(\mathscr{A}) \le \mathbf{Adv}^{(X_1,...,X_\ell)\text{-}\mathrm{mspr}^*}_{\mathsf{TrDM}_c^{\mathsf{P}}}(\mathscr{B}_{\mathrm{mspr}^*}) + \mathbf{Adv}^{IV\text{-}\mathrm{pre}}_{\mathsf{TrDM}_c^{\mathsf{P}}}(\mathscr{B}_{\mathrm{pre}}) \tag{22}$$

$$= \mathbf{Adv}^{(R_1,...,R_\ell)\text{-}\mathrm{mci}}_{\mathsf{P}}(\mathscr{B}_{\mathrm{mspr}^*}) + \mathbf{Adv}^{R_{\mathrm{pre}}(IV)\text{-}\mathrm{ci}}_{\mathsf{P}}(\mathscr{B}_{\mathrm{pre}}). \tag{23}$$

As will be seen, this complicated form using three adversaries is necessary for a fine-grained bound in the setting where $c > h$.

*Proof.* The procedure $\textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}(M)$ evaluates $\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}(M)$ to have the underlying calls $\mathsf{P}(X_1) = Y_1, ..., \mathsf{P}(X_\ell) = Y_\ell$ and outputs $X_1, ..., X_\ell$.

Suppose that $\mathscr{A}(M)$ outputs $M'$ for $\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}(M') = \mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}(M)$. Let $\mathsf{P}(X'_1) = Y'_1, ..., \mathsf{P}(X'_{\ell_2}) = Y'_{\ell_2}$ be the calls in $\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}(M')$. If there exists $j \in \{0, ..., \min\{\ell, \ell_2\} - 1\}$ such that $X_{\ell-j} \neq X'_{\ell_2-j}$, then by outputting $X'_{\ell_2-j}$ for the smallest such $j$, $\mathscr{B}_{\mathrm{mspr}^*}$ finds a second preimage for its $\ell - j$ th challenge input $X_{\ell-j}$. Otherwise, if $\ell > \ell_2$ then $X_{\ell-\ell_2+1} = X'_1 = \star\|IV$, contradicting that $M$ is free-of-IV-preimage; if $\ell < \ell_2$ then $X'_{\ell_2-\ell+1} = X_1 = \star\|IV$, and $\mathscr{B}_{\mathrm{pre}}$ succeeds by outputting $X'_{\ell_2-\ell}$. For rigorousness, we provide pseudocode descriptions of $\mathscr{B}_{\mathrm{mspr}^*}$ and $\mathscr{B}_{\mathrm{pre}}$ in App. E.4. These establish Eq. (22), which implies Eq. (23) by Lemma 1. Similarly to Theorem 2, it also holds in the quantum setting. □

**Corollaries in the plain model.** Lemmas 4 and 2 on preimage and collision security reductions have straightforward corollaries in the plain model.

**Theorem 2.** *Let* $\mathsf{P}$ *be a keyless permutation,* $\mathsf{pd}$ *be an injective padding, and let* $\mathscr{A}$ *be a* $(t, s, \ell_{\max})$*-bounded* $\mathsf{X}$ *adversary.*

- *For any* $Z \in \{0,1\}^h$*, if* $\mathscr{A}$ *breaks the* $Z$*-preimage security of* $\mathsf{Sponge}\text{-}\mathsf{F}_h^{\mathsf{P},\mathsf{pd}}$*, then we can construct a* $(t + O(\ell_{\max}), s + O(\ell_{\max}))$*-bounded* $\mathsf{X}$ *adversary* $\mathscr{B}$ *breaking* $Z$*-preimage security of* $\mathsf{TrDM}_h^{\mathsf{P}}$ *or* $R_{\mathrm{pre}}(Z)$-ci *of* $\mathsf{P}$*, such that*

$$\mathbf{Adv}^{Z\text{-pre}}_{\mathsf{Sponge}\text{-}\mathsf{F}_h^{\mathsf{P},\mathsf{pd}}}(\mathscr{A}) \le \mathbf{Adv}^{Z\text{-pre}}_{\mathsf{TrDM}_h^{\mathsf{P}}}(\mathscr{B}) = \mathbf{Adv}^{R_{\mathrm{pre}}(Z)\text{-ci}}_{\mathsf{P}}(\mathscr{B}). \tag{24}$$

- *If* $\mathscr{A}$ *breaks the collision security of* $\mathsf{Sponge}\text{-}\mathsf{F}_h^{\mathsf{P},\mathsf{pd}}$*, then we can construct a* $(t + O(\ell_{\max}), s + O(\ell_{\max}))$-$\mathsf{X}$ *collision adversary* $\mathscr{B}_{\mathrm{coll}}$ *breaking collision security of* $\mathsf{TrDM}_h^{\mathsf{P}}$ *or* $R_{\mathrm{coll}}$-ci *of* $\mathsf{P}$ *and a* $(t+O(\ell_{\max}), s+O(\ell_{\max}))$*-preimage adversary* $\mathscr{B}_{\mathrm{pre}}$ *breaking* $IV$*-preimage security of* $\mathsf{TrDM}_c^{\mathsf{P}}$ *or* $R_{\mathrm{pre}}(IV)$-ci *of* $\mathsf{P}$*, such that:*

$$\mathbf{Adv}^{\mathrm{coll}}_{\mathsf{Sponge}\text{-}\mathsf{F}_h^{\mathsf{P},\mathsf{pd}}}(\mathscr{A}) \le \mathbf{Adv}^{\mathrm{coll}}_{\mathsf{TrDM}_h^{\mathsf{P}}}(\mathscr{B}_{\mathrm{coll}}) + \mathbf{Adv}^{IV\text{-pre}}_{\mathsf{TrDM}_c^{\mathsf{P}}}(\mathscr{B}_{\mathrm{pre}}) \tag{25}$$

$$= \mathbf{Adv}^{R_{\mathrm{coll}}\text{-ci}}_{\mathsf{P}}(\mathscr{B}_{\mathrm{coll}}) + \mathbf{Adv}^{R_{\mathrm{pre}}(IV)\text{-ci}}_{\mathsf{P}}(\mathscr{B}_{\mathrm{pre}}). \tag{26}$$

We'd like to further expand a bit on the interpretation of Eq. (24). By this:

- If a protocol relies on, e.g., the $0^h$-preimage security of $\mathsf{Sponge}\text{-}\mathsf{F}_h^{\mathsf{P},\mathsf{pd}}$, then one can focus on testifying the (much simpler) $0^h$-preimage security of $\mathsf{TrDM}_h^{\mathsf{P}}$ or the $R_{\mathrm{pre}}(0^h)$-ci of $\mathsf{P}$ via cryptanalysis. As shown in App. A, the standard sponge does not enjoy such a simplification in security evaluation.
- Given a distribution $\mathbf{d}$ over $\{0,1\}^h$, if $\mathbf{Adv}^{\mathbf{d}\text{-}\mathcal{R}_{\mathrm{pre}}\text{-ci}}_{\mathsf{P}}(\mathscr{B}) \le \delta$ for all $(t,s)$-$\mathsf{X}$ adversary $\mathscr{B}$ (i.e., $\mathsf{P}$ is $\mathbf{d}$-oriented CI), then $\mathbf{Adv}^{\mathbf{d}\text{-pre}}_{\mathsf{Sponge}\text{-}\mathsf{F}_h^{\mathsf{P},\mathsf{pd}}}(\mathscr{A}) \le \delta$ for all $(t,s)$-$\mathsf{X}$ adversary $\mathscr{A}$. Such a hash $\mathsf{Sponge}\text{-}\mathsf{F}_h^{\mathsf{P},\mathsf{pd}}$ can be used in protocols that relies on the $Z$-preimage security for a challenge image $Z$ sampled according to $\mathbf{d}$. Furthermore, note that if $\mathbf{d}$ has a high min-entropy then $\delta$ can be small even for non-uniform adversaries.

Regarding second preimage, the major term is $\mathbf{Adv}^{(X_1,\ldots,X_\ell)\text{-mspr}^*}_{\mathsf{TrDM}_c^{\mathsf{P}}}(\mathscr{B})$. Let $bl \in \mathbb{N}$ be an integer and $\ell = \max_{M\in\{0,1\}^{bl}} \left\{|\mathsf{pd}(M)|/r\right\}$ be the maximal number of blocks in padded $bl$-bit messages. Given a subset $\mathcal{S}_1 \subset \{0,1\}^{bl}$ of challenge messages, we can apply $\textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}$ to derive a subset $\mathcal{S}_2 \subset (\{0,1\}^b)^\ell$ of $\ell$-tuples of distinct challenges such that if $\mathbf{Adv}^{(X_1,\ldots,X_\ell)\text{-mspr}^*}_{\mathsf{TrDM}_c^{\mathsf{P}}}(\mathscr{B})$ is small for all $(X_1,\ldots,X_\ell) \in \mathcal{S}_2$ then $\mathbf{Adv}^{M\text{-spr}}_{\mathsf{Sponge}\text{-}\mathsf{F}_h^{\mathsf{P},\mathsf{pd}}}(\mathscr{A})$ is small for all $M \in \mathcal{S}_1$. However, the set $\mathcal{S}_2$ is somewhat contrived. We thereby focus on the distribution-oriented second preimage resistance notion.

**Theorem 3.** *Let* $\mathsf{P}$ *be a keyless permutation,* $\mathsf{pd}$ *be an injective padding,* $bl \in \mathbb{N}$ *be an integer,* $\ell = \max_{M\in\{0,1\}^{bl}} \left\{|\mathsf{pd}(M)|/r\right\}$ *be the maximal number of blocks in padded* $bl$*-bit messages,* $\mathbf{d}$ *be a distribution over* $\{0,1\}^{bl}$ *that is* $(t_D, s_D)$*-samplable by classical uniform algorithms, and let* $\mathscr{A}$ *be a* $(t, s, \ell_{\max})$*-bounded* $\mathsf{X}$ *adversary against the* $\mathbf{d}$*-oriented second preimage security of* $\mathsf{Sponge}\text{-}\mathsf{F}_h^{\mathsf{P},\mathsf{pd}}$*. Then, we can construct:*

*(i)* *A* $(t_D + O(\ell_{\max}), s_D + O(\ell_{\max}))$*-bounded classical adversary* $\mathscr{B}_{\mathrm{coll}}$ *against the collision security of* $\mathsf{TrDM}_h^{\mathsf{P}}$*;*
*(ii)* *A* $(t_D + O(\ell_{\max}), s_D + O(\ell_{\max}))$*-bounded* $\mathsf{X}$ *adversary* $\mathscr{B}_{\mathrm{pre}}$ *against the* $IV$*-preimage security of* $\mathsf{TrDM}_c^{\mathsf{P}}$*;*
*(iii)* *A distribution* $\mathbf{d}_2$ *over* $(\{0,1\}^b)^\ell$ *that is* $(t_D + O(\ell_{\max}), s_D + O(\ell_{\max}))$*-samplable by classical uniform algorithms, and a* $(t + O(\ell_{\max}), s + O(\ell_{\max}))$*-bounded* $\mathsf{X}$ *multi-target second preimage adversary* $\mathscr{B}_{\mathrm{mspr}^*}$ *against* $\mathsf{TrDM}^{\mathsf{P}}$ *for the challenges* $(X_1,\ldots,X_\ell)$*.*

*And the following holds:*

$$\begin{aligned}
&\mathbf{Adv}^{\mathbf{d}\text{-spr}}_{\mathsf{Sponge}\text{-}\mathsf{F}_h^{\mathsf{P},\mathsf{pd}}}(\mathscr{A}) \\
&\le \mathbf{Adv}^{\mathrm{coll}}_{\mathsf{TrDM}_h^{\mathsf{P}}}(\mathscr{B}_{\mathrm{coll}}) + \mathbf{Adv}^{IV\text{-pre}}_{\mathsf{TrDM}_c^{\mathsf{P}}}(\mathscr{B}_{\mathrm{pre}}) + \mathbf{Adv}^{\mathbf{d}_2\text{-mspr}^*}_{\mathsf{TrDM}_c^{\mathsf{P}}}(\mathscr{B}_{\mathrm{mspr}^*}) && (27)\\
&= \mathbf{Adv}^{R_{\mathrm{coll}}\text{-ci}}_{\mathsf{P}}(\mathscr{B}_{\mathrm{coll}}) + \mathbf{Adv}^{R_{\mathrm{pre}}(IV)\text{-ci}}_{\mathsf{P}}(\mathscr{B}_{\mathrm{pre}}) + \mathbf{Adv}^{\mathbf{d}_2\text{-mspr}^*}_{\mathsf{P}}(\mathscr{B}_{\mathrm{mspr}^*}). && (28)
\end{aligned}$$

*Interpretations.* It is natural to ask what does the complicated Theorem 3 tell us. Our opinion is that security now becomes *testifiable*. Concretely, given any approach to generate the second preimage challenge message $M$ for $\mathsf{Sponge}\text{-}\mathsf{F}_h^{\mathsf{P},\mathsf{pd}}$, Theorem 3 describes a corresponding approach to generate a sequence of second preimage challenges $X_1, ..., X_\ell$ for $\mathsf{TrDM}^{\mathsf{P}}$ (and further for the CI of $\mathsf{P}$). It is feasible to evaluate the concrete security margin of $\mathsf{P}$ w.r.t. such challenges using known cryptanalytic tools, and due to Theorem 3, the result directly translates into security margin of $\mathsf{Sponge}\text{-}\mathsf{F}_h^{\mathsf{P},\mathsf{pd}}$. The proof is given in App. E.5.

### 4.3 Precise Bounds In the (Quantum) Random Permutation Model

Lemma 2 and 3 also enable deriving concrete security bounds in the (quantum) random permutation model. Due to using (quantum) random permutation, we can now prove everywhere (second) preimage resistance.

**Theorem 4.** *Assume $b \geq 3$. Consider the construction $\mathsf{Sponge}\text{-}\mathsf{F}_h^{\mathbf{\Pi},\mathsf{pd}}$ using a $b$-bit random permutation $\mathbf{\Pi}$, output length $\nu \in \{h, ..., \nu_{\max}\}$, capacity $c \geq h$ and rate $r$. Then: for any $(q, t, s, \ell_{\max})$-bounded classical adversary $\mathscr{A}^{\mathbf{\Pi},\mathbf{\Pi}^{-1}}$ with $q+2\ell_{\max} \leq 2^b/2$, $\mathsf{Sponge}\text{-}\mathsf{F}_h^{\mathbf{\Pi},\mathsf{pd}}$ has collision, everywhere preimage and everywhere second preimage security bounds as follows:*

$$\mathbf{Adv}^{\text{coll}}_{\mathsf{Sponge}\text{-}\mathsf{F}_h^{\mathbf{\Pi},\mathsf{pd}}}(\mathscr{A}) \leq \frac{(q+2\ell_{\max})^2}{2^h} + \frac{2(q+2\ell_{\max})}{2^h}; \tag{29}$$

$$\mathbf{Adv}^{\text{epre}}_{\mathsf{Sponge}\text{-}\mathsf{F}_h^{\mathbf{\Pi},\mathsf{pd}}}(\mathscr{A}) \leq \frac{2(q+\ell_{\max})}{2^h}; \tag{30}$$

$$\mathbf{Adv}^{\text{espr}[bl]}_{\mathsf{Sponge}\text{-}\mathsf{F}_h^{\mathbf{\Pi},\mathsf{pd}}}(\mathscr{A}) \leq \frac{\ell^2}{2^h} + \frac{2\ell}{2^c} + \frac{2\ell(q+2\ell_{\max})}{2^c} + \frac{2(q+2\ell_{\max})}{2^h}. \tag{31}$$

*where $\ell = \max_{M\in\{0,1\}^{bl}} \left\{|\mathsf{pd}(M)|/r\right\}$ is the maximal number of blocks in padded challenges messages.*

*For any $(q, t, s, \ell_{\max})$-bounded quantum adversary $\mathscr{A}^{\mathbf{\Pi},\mathbf{\Pi}^{-1}}$, $\mathsf{Sponge}\text{-}\mathsf{F}_h^{\mathbf{\Pi},\mathsf{pd}}$ has quantum collision, quantum everywhere preimage and quantum everywhere second preimage security bounds as follows:*

$$\mathbf{Adv}^{\text{q-coll}}_{\mathsf{Sponge}\text{-}\mathsf{F}_h^{\mathbf{\Pi},\mathsf{pd}}}(\mathscr{A}) \leq \frac{4(8q+16\ell_{\max}+1)^4}{2^h} + \frac{2(8q+16\ell_{\max}+1)^2}{2^c}; \tag{32}$$

$$\mathbf{Adv}^{\text{q-epre}}_{\mathsf{Sponge}\text{-}\mathsf{F}_h^{\mathbf{\Pi},\mathsf{pd}}}(\mathscr{A}) \leq \frac{2(8q+8\ell_{\max}+1)^2}{2^h}; \tag{33}$$

$$\mathbf{Adv}^{\text{q-espr}[bl]}_{\mathsf{Sponge}\text{-}\mathsf{F}_h^{\mathbf{\Pi},\mathsf{pd}}}(\mathscr{A}) \leq \frac{\ell^2}{2^h} + \frac{2\ell}{2^c} + \frac{2\ell(8q+16\ell_{\max}+1)^2}{2^c} + \frac{2(8q+16\ell_{\max}+1)^2}{2^h}. \tag{34}$$

*Proof.* We prove the claims in turn.

*Collision.* By Lemma 2, it suffices to bound $\mathbf{Adv}^{R_{\text{coll}}\text{-ci}}_{\mathbf{\Pi}}(\mathscr{B}_{\text{coll}})$ for $(q+2\ell_{\max}, t+O(\ell_{\max}), s+O(\ell_{\max}))$-bounded $\mathscr{B}_{\text{coll}}$ and $\mathbf{Adv}^{R_{\text{pre}}(IV)\text{-ci}}_{\mathbf{\Pi}}(\mathscr{B}_{\text{pre}})$ for $(q+2\ell_{\max}, t+O(\ell_{\max}), s+O(\ell_{\max}))$-bounded $\mathscr{B}^{\mathbf{\Pi},\mathbf{\Pi}^{-1}}_{\text{pre}}$. Consider every pair of $\mathscr{B}$'s $\mathbf{\Pi}$-queries $\mathbf{\Pi}(X) = Y$ and $\mathbf{\Pi}(X') = Y'$, where $\mathbf{\Pi}(X') = Y'$ appears after $\mathbf{\Pi}(X) = Y$. Then, regardless of direction of $\mathbf{\Pi}(X') = Y'$, $X'\oplus Y'$ is uniform in a set of size at least $2^b - q - 2\ell_{\max}$. Among them, the number of choices of $X'\oplus Y'$ with $\mathsf{right}_h(X'\oplus Y') = \mathsf{right}_h(X\oplus Y)$ is at most $2^{b-h}$. Thus, $\Pr\left[\mathsf{right}_h(X'\oplus Y') = \mathsf{right}_h(X\oplus Y)\right] \leq \frac{2^{b-h}}{2^b-q-2\ell_{\max}} \leq \frac{2}{2^h}$ (since $q+2\ell_{\max} \leq 2^b/2 \Rightarrow 2^b - 2q - 2\ell_{\max} \leq 2^b/2$). Summing over at most $\binom{q+2\ell_{\max}}{2}$ choices of $\mathbf{\Pi}(X) = Y, \mathbf{\Pi}(X') = Y'$ yields

$$\mathbf{Adv}^{R_{\text{coll}}\text{-ci}}_{\mathbf{\Pi}}(\mathscr{B}) \leq \binom{q+2\ell_{\max}}{2} \times \frac{2}{2^h} \leq \frac{(q+2\ell_{\max})^2}{2^h} + \frac{2(q+2\ell_{\max})}{2^h}. \tag{35}$$

Similarly, every $\mathbf{\Pi}$-query $\mathbf{\Pi}(X) = Y$ of $\mathscr{B}^{\mathbf{\Pi},\mathbf{\Pi}^{-1}}_{\text{pre}}$ has $\Pr\left[\mathsf{right}_h(X\oplus Y) = \mathsf{left}_h(Z)\right] \leq \frac{2^{b-h}}{2^b-q-\ell_{\max}} \leq \frac{2}{2^h}$. Summing over $q+2\ell_{\max}$ queries of $\mathscr{B}^{\mathbf{\Pi},\mathbf{\Pi}^{-1}}_{\text{pre}}$ yields

$$\mathbf{Adv}^{R_{\text{pre}}(IV)\text{-ci}}_{\mathbf{\Pi}}(\mathscr{B}_{\text{pre}}) \leq \frac{2(q+2\ell_{\max})}{2^h}. \tag{36}$$

Gathering the two bounds establishes Eq. (29).

In the quantum setting, we apply Theorem 1: since $R_{\text{coll}}$ is a 4-ary relation, its parameter $k$ for Theorem 1 is 2, and for any $q+2\ell_{\max}$ query quantum adversary $\mathscr{B}^{\mathbf{\Pi},\mathbf{\Pi}^{-1}}$ there exists a 2-query classical adversary $\mathscr{A}^{\mathbf{\Pi},\mathbf{\Pi}^{-1}}$ such that

$$\frac{\left(1-\frac{4}{2^b}\right)}{(8q+16\ell_{\max}+1)^4} \cdot \mathbf{Adv}^{\text{q-}R_{\text{coll}}\text{-ci}}_{\mathbf{\Pi}}(\mathscr{B}) \leq \mathbf{Adv}^{R_{\text{coll}}\text{-ci}}_{\mathbf{\Pi}}(\mathscr{A}). \tag{37}$$

A single query pair $\boldsymbol{\Pi}(X) = Y, \boldsymbol{\Pi}(X') = Y'$ yields $\mathsf{right}_h(Y \oplus X) = \mathsf{right}_h(Y' \oplus X')$ with probability $\leq 2^{b-h}/2^{b-1} \leq 2/2^h$. Moreover, $b \geq 3 \Rightarrow 1 - \frac{4}{2^b} \geq \frac{1}{2}$. Therefore,

$$\mathbf{Adv}^{\text{q-}R_{\text{coll}}\text{-ci}}_{\boldsymbol{\Pi}}(\mathscr{B}) \leq \frac{2}{2^h} \times \frac{(8q + 16\ell_{\max} + 1)^4}{\left(1 - \frac{4}{2^b}\right)} \leq \frac{4(8q + 16\ell_{\max} + 1)^4}{2^h}. \tag{38}$$

Similarly, for any $q + 2\ell_{\max}$ query quantum adversary $\mathscr{A}^{\boldsymbol{\Pi},\boldsymbol{\Pi}^{-1}}$ there exists a 1-query classical adversary $\mathscr{B}^{\boldsymbol{\Pi},\boldsymbol{\Pi}^{-1}}$ such that

$$\frac{\left(1 - \frac{1}{2^b}\right)}{(8q + 16\ell_{\max} + 1)^2} \cdot \mathbf{Adv}^{\text{q-}R_{\text{pre}}(IV)\text{-ci}}_{\boldsymbol{\Pi}}(\mathscr{A}) \leq \mathbf{Adv}^{R_{\text{pre}}(IV)\text{-ci}}_{\boldsymbol{\Pi}}(\mathscr{B}). \tag{39}$$

The single pair of queries of $\mathscr{B}$ yields $(X, Y) : \mathsf{right}_c(X \oplus Y) = \mathsf{left}_c(Z)$ with probability exactly $1/2^c$. Moreover, $b \geq 3 \Rightarrow 1 - \frac{1}{2^b} \geq \frac{1}{2}$. Therefore,

$$\mathbf{Adv}^{\text{q-}R_{\text{pre}}(IV)\text{-ci}}_{\boldsymbol{\Pi}}(\mathscr{A}) \leq \frac{1}{2^c} \times \frac{(8q + 16\ell_{\max} + 1)^2}{\left(1 - \frac{1}{2^b}\right)} \leq \frac{2(8q + 16\ell_{\max} + 1)^2}{2^c}. \tag{40}$$

Gathering the two yields Eq. (32).

*Everywhere preimage.* We consider $\mathbf{Adv}^{R_{\text{pre}}(Z)\text{-ci}}_{\boldsymbol{\Pi}}(\mathscr{B}_{\text{pre}})$ for $(q+\ell_{\max}, t+O(\ell_{\max}), s+O(\ell_{\max}))$-bounded $\mathscr{B}^{\boldsymbol{\Pi},\boldsymbol{\Pi}^{-1}}_{\text{pre}}$. Injecting the query complexity $q+\ell_{\max}$ and the challenge image size $h$ into the above analyses for $\mathbf{Adv}^{R_{\text{pre}}(IV)\text{-ci}}_{\boldsymbol{\Pi}}$ and $\mathbf{Adv}^{\text{q-}R_{\text{pre}}(IV)\text{-ci}}_{\boldsymbol{\Pi}}$ yields Eq. (30) and (33).

*Everywhere second preimage.* This requires bounding the $M$-spr advantage for any $M$, over the random choice of $\boldsymbol{\Pi}$. Let $\mathsf{BadM}$ be the event that $M$ is *not* free-of-inner-collision or *not* free-of-IV-preimage. Then, over the random choice of $\boldsymbol{\Pi}$, the probability of $\mathsf{BadM}$ is at most $\binom{\ell}{2} \cdot \frac{2^{b-h}}{2^b - \ell} + \frac{\ell 2^{b-c}}{2^b - \ell} \leq \frac{\ell^2}{2^h} + \frac{2\ell}{2^c}$, which is sufficiently small.

By Lemma 3, for any free-of-inner-collision and free-of-IV-preimage $M$ and any $(q, t, s, \ell_{\max})$-bounded adversary $\mathscr{A}^{\boldsymbol{\Pi},\boldsymbol{\Pi}^{-1}}$ against $M$-spr of $\mathsf{Sponge}\text{-}\mathsf{F}^{\mathsf{P},\mathsf{pd}}$, we can construct $\ell$ distinct challenges $X_1, ..., X_\ell$ and:

(i) A $(q+2\ell_{\max}, t+O(\ell_{\max}), s+O(\ell_{\max}))$-bounded $\mathscr{B}^{\boldsymbol{\Pi},\boldsymbol{\Pi}^{-1}}_{\text{mspr}^*}$ against security $(R_{\text{spr}}(X_1, c), ..., R_{\text{spr}}(X_{\ell-1}, c), R_{\text{spr}}(X_\ell, h))$-mci of $\boldsymbol{\Pi}$;
(ii) A $(q + 2\ell_{\max}, t + O(\ell_{\max}), s + O(\ell_{\max}))$-bounded $\mathscr{B}^{\boldsymbol{\Pi},\boldsymbol{\Pi}^{-1}}_{\text{pre}}$ for $R_{\text{pre}}(IV)$-ci of the $b$-bit random permutation $\boldsymbol{\Pi}$.

If the above are small for *all "free-of-inner-collision" tchallenges* $(X_1, ..., X_\ell)$ (which is possible in the random permutation model), then $\mathbf{Adv}^{M\text{-spr}}_{\mathsf{Sponge}\text{-}\mathsf{F}^{\mathsf{P},\mathsf{pd}}}(\mathscr{A})$ is small for all free-of-inner-collision and free-of-IV-preimage message $M$ such that $\mathsf{pd}(M)$ has $\ell$ blocks.

As proved before, $\mathbf{Adv}^{R_{\text{pre}}(IV)\text{-ci}}_{\boldsymbol{\Pi}}(\mathscr{B}_{\text{pre}}) \leq \frac{2(q+2\ell_{\max})}{2^c}$. Let

$$\delta_1 := \max_{X' \in \{0,1\}^b} \left\{ \mathbf{Adv}^{R_{\text{spr}}(X',c)\text{-ci}}_{\boldsymbol{\Pi}}(\mathscr{B}) \right\}, \quad \delta_2 := \max_{X' \in \{0,1\}^b} \left\{ \mathbf{Adv}^{R_{\text{spr}}(X',h)\text{-ci}}_{\boldsymbol{\Pi}}(\mathscr{B}) \right\}. \tag{41}$$

We show that

$$\max_{X_1,...,X_\ell} \left\{ \mathbf{Adv}^{(R_{\text{spr}}(X_1,c),...,R_{\text{spr}}(X_{\ell-1},c),R_{\text{spr}}(X_\ell,h))\text{-mci}}_{\boldsymbol{\Pi}}(\mathscr{B}) \right\} \leq (\ell - 1)\delta_1 + \delta_2. \tag{42}$$

For this, assume

$$\max_{X_1,...,X_\ell} \left\{ \mathbf{Adv}^{(R_{\text{spr}}(X_1,c),...,R_{\text{spr}}(X_{\ell-1},c),R_{\text{spr}}(X_\ell,h))\text{-mci}}_{\boldsymbol{\Pi}}(\mathscr{B}) \right\} > (\ell - 1)\delta_1 + \delta_2$$

towards a contradiction. Then,

$$\begin{aligned}
&\mathbf{Adv}^{(R_{\text{spr}}(X_1,c),...,R_{\text{spr}}(X_\ell,h))\text{-mci}}_{\boldsymbol{\Pi}}(\mathscr{B}) \\
&= \Pr\left[\mathscr{B}^{\boldsymbol{\Pi},\boldsymbol{\Pi}^{-1}} \Rightarrow X' : X' \in R_{\text{spr}}(X_\ell, h)\right] + \sum_{i=1}^{\ell-1} \Pr\left[\mathscr{B}^{\boldsymbol{\Pi},\boldsymbol{\Pi}^{-1}} \Rightarrow X' : X' \in R_{\text{spr}}(X_i, c)\right].
\end{aligned}$$

As $\Pr\left[\mathscr{B}^{\boldsymbol{\Pi},\boldsymbol{\Pi}^{-1}} \Rightarrow X' : X' \in R_{\text{spr}}(X_\ell, h)\right] \leq \delta_2$ by Eq. (41), $\sum_{i=1}^{\ell-1} \Pr\left[\mathscr{B}^{\boldsymbol{\Pi},\boldsymbol{\Pi}^{-1}} \Rightarrow X' : X' \in R_{\text{spr}}(X_i, c)\right] > (\ell - 1)\delta_1$. By the pigeonhole principle, there exists $X_i \in \{X_1, ..., X_{\ell-1}\}$ such that $\Pr\left[\mathscr{B}^{\boldsymbol{\Pi},\boldsymbol{\Pi}^{-1}} \Rightarrow X' : X' \in R_{\text{spr}}(X_i, c)\right] = \mathbf{Adv}^{R_{\text{spr}}(X_i,c)\text{-ci}}_{\boldsymbol{\Pi}}(\mathscr{B}) > \delta_1$, contradicting Eq. (41). These establish Eq. (42).

It then suffices to bound $\mathbf{Adv}_{\mathbf{\Pi}}^{R_{\rm spr}(X',c)\text{-ci}}(\mathscr{B})$ and $\mathbf{Adv}_{\mathbf{\Pi}}^{R_{\rm spr}(X',h)\text{-ci}}(\mathscr{B})$ for any $X' \in \{0,1\}^b$. Consider every $\mathbf{\Pi}$-query of $\mathscr{B}^{\mathbf{\Pi},\mathbf{\Pi}^{-1}}$: as argued, regardless of whether it is $\mathbf{\Pi}(X) \to Y$ or $\mathbf{\Pi}^{-1}(Y) \to X$, the resulted $X \oplus Y$ is uniform in at least $2^b - q - 2\ell_{\max}$ possibilities, the probability to have $\mathsf{right}_c(X \oplus Y) = \mathsf{right}_c(X' \oplus \mathbf{\Pi}(X'))$ is at most $\frac{2^{b-c}}{2^b - q - 2\ell_{\max}} \le \frac{2}{2^c}$ (since $q + 2\ell_{\max} \le 2^b/2$). Summing over the at most $q + 2\ell_{\max}$ queries of $\mathscr{B}^{\mathbf{\Pi},\mathbf{\Pi}^{-1}}$ yields $\delta_1 \le \frac{2(q+2\ell_{\max})}{2^c}$ for the quantity defined in Eq. (41). In a similar vein, it can be shown $\delta_2 \le \frac{2(q+2\ell_{\max})}{2^h}$.

It thus holds

$$\begin{aligned}\mathbf{Adv}^{M\text{-spr}}_{\mathsf{Sponge}\text{-}\mathsf{F}^{\mathsf{p},\mathsf{pd}}}(\mathscr{A}) &\le \Pr\big[\mathsf{BadM}\big] + \underbrace{\Pr\big[\mathscr{B}_{\rm pre} \text{ succeeds} \mid \neg\mathsf{BadM}\big]}_{\le \mathbf{Adv}_{\mathbf{\Pi}}^{R_{\rm pre}(IV)\text{-mci}}(\mathscr{B}_{\rm pre}) \le \frac{2(q+2\ell_{\max})}{2^c}} \\ &\quad + \underbrace{\Pr\big[\mathscr{B}_{\rm mspr^*} \text{ succeeds} \mid \neg\mathsf{BadM}\big]}_{\le \mathbf{Adv}_{\mathbf{\Pi}}^{(R_{\rm spr}(X_1,c),\ldots,R_{\rm spr}(X_\ell,h))\text{-mci}}(\mathscr{B}) \le (\ell-1)\delta_1 + \delta_2} \\ &\le \frac{\ell^2}{2^h} + \frac{2\ell}{2^c} + \frac{2\ell(q+2\ell_{\max})}{2^c} + \frac{2(q+2\ell_{\max})}{2^h}.\end{aligned} \tag{43}$$

Since this holds for any $M$ such that $\mathsf{pd}(M)$ has $\ell$ blocks, Eq. (30) is established.

In the quantum setting, the expansion in Eq. (43) still holds, and we apply Theorem 1 to bound $\mathbf{Adv}_{\mathbf{\Pi}}^{\text{q-}R_{\rm spr}(X',c)\text{-ci}}(\mathscr{A})$ and $\mathbf{Adv}_{\mathbf{\Pi}}^{\text{q-}R_{\rm spr}(X',h)\text{-ci}}(\mathscr{A})$ for any $X' \in \{0,1\}^b$ and any *quantum* $(q+2\ell_{\max}, t+O(\ell_{\max}), s+O(\ell_{\max}))$-adversary $\mathscr{B}$. The relation $R_{\rm spr}(X',c)$ is also 2-ary, meaning that for any $q + 2\ell_{\max}$ query quantum adversary $\mathscr{B}^{\mathbf{\Pi},\mathbf{\Pi}^{-1}}$ there exists a 1-query classical adversary $\mathscr{A}^{\mathbf{\Pi},\mathbf{\Pi}^{-1}}$ such that

$$\frac{\left(1 - \frac{1}{2^b}\right)}{(8q + 16\ell_{\max} + 1)^2} \cdot \mathbf{Adv}_{\mathbf{\Pi}}^{\text{q-}R_{\rm spr}(X',c)\text{-ci}}(\mathscr{B}) \le \mathbf{Adv}_{\mathbf{\Pi}}^{R_{\rm spr}(X',c)\text{-ci}}(\mathscr{A}). \tag{44}$$

The single pair of queries of $\mathscr{A}$ yields $(X,Y) : \mathsf{right}_c(Y \oplus X) = \mathsf{right}_c(\mathbf{\Pi}(X') \oplus X')$ with probability at most $1/2^c$. Moreover, when $b \ge 3$ it holds $1 - \frac{1}{2^b} \ge \frac{1}{2}$. Therefore,

$$\mathbf{Adv}_{\mathbf{\Pi}}^{\text{q-}R_{\rm spr}(X',c)\text{-ci}}(\mathscr{B}) \le \frac{1}{2^c} \times \frac{(8q + 16\ell_{\max} + 1)^2}{\left(1 - \frac{1}{2^b}\right)} \le \frac{2(8q + 16\ell_{\max} + 1)^2}{2^c}. \tag{45}$$

Similarly,

$$\mathbf{Adv}_{\mathbf{\Pi}}^{\text{q-}R_{\rm spr}(X',h)\text{-ci}}(\mathscr{B}) \le \frac{1}{2^h} \times \frac{(8q + 16\ell_{\max} + 1)^2}{\left(1 - \frac{1}{2^b}\right)} \le \frac{2(8q + 16\ell_{\max} + 1)^2}{2^h}. \tag{46}$$

On the other hand, $\mathbf{Adv}_{\mathbf{\Pi}}^{\text{q-}R_{\rm pre}(IV)\text{-ci}}(\mathscr{B})$ has been bounded to $\frac{2(8q+16\ell_{\max}+1)^2}{2^c}$ in Eq. (40). Injecting these into Eq. (43) yields Eq. (33). □

Recall from Sect. 2.2 that a relation $R$ is (quantum) evasive in the asymptotic sense, if $\mathbf{Adv}_{\mathbf{\Pi}}^{R\text{-ci}}(\mathscr{A})$ is negligible for every efficient (quantum) adversary $\mathscr{A}$. Hence, Eqs. (35) and (36), resp. (38), (40), (45) and (46) have established evasiveness, resp. quantum evasiveness, for the involved relations.

## 5 Security of Sponge-$\mathsf{F}^{\mathbf{\Pi},\mathsf{pd}}$ for LMS Signature

Leighton-Micali Signature (LMS) [61] is a stateful signature constructed by combining a variant of Winternitz one-time signature [55,69,68] and several Merkle-trees [69,68]. Both components are built from a hash function $\mathsf{h}$ that can be chosen from SHA-256 or SHAKE, and have security reductions to certain forms of preimage security of $\mathsf{h}$. LMS has been published as RFC 8554 [64] and included in NIST SP 800-208 [29].

The provable security of LMS of [39, Corollary 1] relies on a multi-target fixed-prefix preimage security of the hash function $\mathsf{h}$, which is not implied by standard second preimage security. Below we first formalize this notion, and then prove security for Sponge-$\mathsf{F}^{\mathbf{\Pi},\mathsf{pd}}$ regarding this notion.

**The** lms-mfpspr **experiment** $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{h}^{\mathbf{\Pi}},\mathscr{A},\mu,\Phi}$

1. **Initialization:** Sample a random permutation $\mathbf{\Pi} \xleftarrow{\$} \mathcal{P}(b)$. Pick three groups of $\mathbf{\Pi}$-independent prefixes $\mathcal{G}_1 = \big(\mathsf{prefix}_{1,1}, ..., \mathsf{prefix}_{1,\gamma_1}\big)$, $\mathcal{G}_2 = \big(\mathsf{prefix}_{2,1}, ..., \mathsf{prefix}_{2,\gamma_2}\big)$, $\mathcal{G}_3 = \big(\mathsf{prefix}_{3,1}, ..., \mathsf{prefix}_{3,\gamma_3}\big)$ in arbitrary, subjecting to constraints as follows:
   (a) $\max_{PP} \big|\big\{(i,j) : \mathsf{prefix}_{i,j} = PP\big\}\big| \leq \mu$;
   (b) There does not exist distinct $(i,j), (i',j')$ such that $\mathsf{prefix}_{i,j} = \mathsf{prefix}_{i',j'} \| \star$;
   (c) $\mathcal{G}_1 \cap \mathcal{G}_2 = \emptyset$, $\mathcal{G}_1 \cap \mathcal{G}_3 = \emptyset$, $\mathcal{G}_2 \cap \mathcal{G}_3 = \emptyset$.
2. **Phase 1 (Generate challenge digests for $\mathcal{G}_1$ and $\mathcal{G}_2$):**
   (a) For each $\mathsf{prefix}_{1,i} \in \mathcal{G}_1$: $M^{1,i} \xleftarrow{\$} \{0,1\}^h$, $Z^{1,i} \leftarrow \mathsf{h}^{\mathbf{\Pi}}(\mathsf{prefix}_{1,i} \| M^{1,i})$
   (b) For each $\mathsf{prefix}_{2,i} \in \mathcal{G}_2$: choose $M^{2,i}$ in arbitrary and compute $Z^{2,i} \leftarrow \mathsf{h}^{\mathbf{\Pi}}(\mathsf{prefix}_{2,i} \| M^{2,i})$
   Set
   $$Inputs \leftarrow \big((Z^{1,1}, ..., Z^{1,\gamma_1}), (M^{2,1}, Z^{2,1}), ..., (M^{2,\gamma_2}, Z^{2,\gamma_2})\big).$$
3. **Phase 2 (Adaptively generate challenges digests for $\mathcal{G}_3$):**
   For $i = 1, ..., \gamma_3$, do
   (a) $M^{3,i} \leftarrow \mathscr{A}^{\mathbf{\Pi},\mathbf{\Pi}^{-1}}\big(St, Inputs\big)$
   (b) $R^i \xleftarrow{\$} \{0,1\}^h$, $Z^{3,i} \leftarrow \mathsf{h}^{\mathbf{\Pi}}(\mathsf{prefix}_{3,i} \| R^i \| M^{3,i})$
   (c) Add $(R^i, Z^{3,i})$ to $Inputs$
4. *Remark*: All the involved values
   $$\big\{\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \{M^{i,j}\}_{i=1,2,3;j=1,...,\gamma_i}, \{R^i\}_{i=1,...,\gamma_3}\big\}$$
   must fulfill the constraints defined in $\Phi$.
5. **Phase 3:** $M' \leftarrow \mathscr{A}^{\mathbf{\Pi},\mathbf{\Pi}^{-1}}\big(St, Inputs\big)$
6. **Finalization:** Output 1, if any of the following is fulfilled:
   (a) There exists $i$ such that $\mathsf{h}^{\mathbf{\Pi}}(\mathsf{prefix}_{1,i} \| M') = Z^{1,i}$;
   (b) There exists $i$ such that $\mathsf{h}^{\mathbf{\Pi}}(\mathsf{prefix}_{2,i} \| M') = Z^{2,i}$, whereas $M' \neq M^{2,i}$;
   (c) There exists $i$ such that $\mathsf{h}^{\mathbf{\Pi}}(\mathsf{prefix}_{3,i} \| M') = Z^{3,i}$, whereas $M' \neq R^i \| M^{3,i}$.

**Fig. 4:** The lms-mfpspr experiment $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{h}^{\mathbf{\Pi}},\mathscr{A},\mu,\Phi}$.

**LMS-Multi-target Fixed-Prefix Preimage Security.** Fluhrer [39] studied LMS in the multi-user unforgeability game, and formalized an experiment in the random oracle model that summarizes all possible inputs and outputs of the underlying hash construction $\mathsf{h}$ in this unforgeability game. We give a pseudocode description of $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{h}^{\mathbf{\Pi}},\mathscr{A},\mu,\Phi}$, a random permutation-based variant of Fluhrer's experiment, in Fig. 4. Roughly speaking, in $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{h}^{\mathbf{\Pi}},\mathscr{A},\mu,\Phi}$, the adversary $\mathscr{A}^{\mathbf{\Pi},\mathbf{\Pi}^{-1}}$ is given a series of strings with prefixes (generated randomly yet in a structured manner) and hash digests. The adversarial goal is to find one more string that has the same prefix and digest as any of its input strings.

**Definition 1 (**lms-mfpspr **advantage).** *Given a hash function $\mathsf{h}^{\mathbf{\Pi}} : \{0,1\}^* \mapsto \{0,1\}^h$, an adversary $\mathscr{A}^{\mathbf{\Pi},\mathbf{\Pi}^{-1}}$, and a set $\Phi$ of constraints on the values involved in the experiment $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{h}^{\mathbf{\Pi}},\mathscr{A},\mu,\Phi}$, the LMS-multi-target, fixed-prefix second preimage attack advantage of $\mathscr{A}^{\mathbf{\Pi},\mathbf{\Pi}^{-1}}$ against $\mathsf{h}^{\mathbf{\Pi}}$ is defined as*

$$\mathbf{Adv}^{\text{lms-mfpspr}[\Phi]}_{\mathsf{h}^{\mathbf{\Pi}}}(\mathscr{A}) := \Pr\big[\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{h}^{\mathbf{\Pi}},\mathscr{A},\mu,\Phi} \Rightarrow 1\big].$$

**lms-mfpspr Security for Sponge-$\mathsf{F}^{\mathbf{\Pi},\mathsf{apd}}$.** We also fix $h$ and view Sponge-$\mathsf{F}^{\mathbf{\Pi},\mathsf{apd}}$ as a fixed-output-length hash Sponge-$\mathsf{F}^{\mathbf{\Pi},\mathsf{apd}}_h(\cdot) :=$ Sponge-$\mathsf{F}^{\mathbf{\Pi},\mathsf{apd}}(\cdot, h)$. To fit into Ascon-p-based instance of Sponge-$\mathsf{F}^{\mathbf{\Pi},\mathsf{apd}}_h$, we consider constraints that are different from Fluhrer [39, Theorem 1].

**Theorem 5.** *Let $\Phi$ be the following constraints:*

*(i) Every prefix $\mathsf{prefix}_{i,j}$ has $|\mathsf{prefix}_{i,j}| < 3r$;*
*(ii) Every prefix $\mathsf{prefix}_{i,j}$ is of the form $I^{\mathsf{idx}} \| \star$ for some $\mathsf{idx} \in \{1, ..., u\}$, where $u$ is an integer determined by LMS, and every $I^{\mathsf{idx}}$ has $|I^{\mathsf{idx}}| = 2r$.*

*Moreover, let $\ell_{\max}$ be the maximal number of blocks in padded messages of* $\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}}$*. Then, when the total number of* $\mathbf{\Pi}$*-queries appeared during the experiment* $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}},\mathscr{A},\mu,\Phi}$ *is* $Q$*,* $4Q^2 \le 2^h$*, for any integer threshold* $C$ *it holds*

$$\mathbf{Adv}^{\text{lms-mfpspr}[\Phi]}_{\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}}}(\mathscr{A}) \le \frac{8u^2 + 4\mu C^2 \ell_{\max} Q + 2^{r+2}(\mu C + 1)Q + 84\mu C h Q}{2^h} \tag{47}$$

$$+ \frac{4Q^3}{2^{2h}} + \Big(\frac{2Q^2}{2^h}\Big)^C \cdot \frac{1}{C!} + \frac{8Q^2}{h!2^h} + \frac{4\mu C^2 \ell_{\max}^2 Q}{2^b}. \tag{48}$$

The proof is given in App. F. It relies on a new bound on the probability of the success condition (6.c) in Fig. 4, which has taken the maximal number of blocks $\ell_{\max}$ in padded messages.

Eq. (48) can be injected into Fluhrer's subsequent analyses of LMS (proof of [39, Theorem 2]), and final bound is Eq. (48) plus $\frac{u^{\mu+1}}{2^{128\mu}} \cdot \frac{1}{(\mu+1)!}$. Ascon-Sp-F has $r = 64 = h/4$. When $u = 2^{64}$, $Q = 2^{120}$ (as considered in [39, Corollary 1]), $\ell_{\max} \le 2^{55} + 1$ (as in SHA-256), $\mu = 3$ and $C = 6$, the dominating terms are $\frac{4\mu C^2 \ell_{\max} Q}{2^h} \le \frac{432 \times 2^{184}}{2^{256}} \le \frac{1}{2^{63}}$ and $\frac{2^{r+2}(\mu C+1)Q}{2^h} \le \frac{19 \times 2^{186}}{2^{256}} \le \frac{1}{2^{65}}$, which is less than our claimed success probability bound $2^{-60}$.

## 6 Indifferentiability of $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$

*Indifferentiability of XOFs.* For indifferentiability, we will view $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ as a (sponge-like) extendable-output function (XOF). According to [45,57] and NIST FIPS 202 [37], such a function $\mathrm{XOF}(M, \nu)$ takes a message $M \in \{0,1\}^*$ and an integer $\nu$ as inputs, maps $M$ to a long string and outputs the first $\nu$ bits of this string. We follow [45,37,57] and take such a XOF as the ideal reference of our construction. Following [45,57], we consider a variable-output-length random oracle $\mathbf{VRO} : \{0,1\}^* \mapsto \{0,1\}^*$, and consider that the random oracle queried with input-pair $(M, \nu) \in \{0,1\}^* \times \{h, ..., \nu_{\max}\}$ gives $\mathsf{left}_\nu\big(\mathbf{VRO}(M)\big)$. Let $\mathcal{S}^{\mathbf{VRO}}$ be a simulator that queries $\mathbf{VRO}$ and provides the same interfaces as $\mathbf{\Pi}$ and $\mathbf{\Pi}^{-1}$. Then, for any distinguisher $\mathcal{D}$, the indifferentiability advantage of $\mathcal{D}$ against $\mathrm{XOF}^{\mathbf{\Pi}}$ is

$$\mathbf{Adv}^{\text{indiff}}_{\mathrm{XOF}^{\mathbf{\Pi}},\mathbf{\Pi},\mathcal{S}}(\mathcal{D}) := \Big|\Pr\big[\mathcal{D}^{\mathbf{\Pi},\mathbf{\Pi}^{-1},\mathrm{XOF}^{\mathbf{\Pi}}} \Rightarrow 1\big] - \Pr\big[\mathcal{D}^{\mathcal{S}^{\mathbf{VRO}},\mathrm{P},\mathcal{S}^{\mathbf{VRO}},\mathrm{P}^{-1},\mathbf{VRO}} \Rightarrow 1\big]\Big|.$$

*Indifferentiability Result of* $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$. Following Bertoni et al. [13], we measure the cost of a construction query based on the number of permutation evaluations required in the real world to produce the output. Concretely, each call to $\mathsf{Sponge\text{-}F}^{\mathbf{\Pi},\mathsf{pd}}(M, \nu)$ with $\ell$-block $\overline{M}$ has a cost of $\ell$. Therefore, if the adversary makes $p$ queries to $\mathbf{\Pi}$ and $\mathbf{\Pi}^{-1}$ and $q$ queries to $\mathsf{Sponge\text{-}F}^{\mathbf{\Pi},\mathsf{pd}}$ and the padded messages have at most $\sigma$ blocks in total (under our parsing role), then the *total oracle query cost* is at most $Q \le p + \sigma$, which is the $Q$ value introduced in Theorem 6. With these, our main claim is as follows.

**Theorem 6.** *Assume* $r' \ge 3$ *and* $Q \ge 8$*. Then, there exists a simulator* $\mathcal{S}^{\mathbf{VRO}}$ *such that the following holds for any differentiator* $\mathcal{D}$ *with total number of oracle query costs* $Q$*:*

$$\mathbf{Adv}^{\text{indiff}}_{\mathsf{Sponge\text{-}F}^{\mathbf{\Pi},\mathsf{pd}},\mathbf{\Pi},\mathcal{S}}(\mathcal{D}) \le \frac{1}{2^{r'}} + \frac{(4r'+2)Q^2}{2^b} + \frac{12Q + 14Q^2}{2^c} + \frac{5Q\log_2 Q}{2^{b-r'}}.$$

*Moreover,* $\mathcal{S}^{\mathbf{VRO}}$ *makes at most* $Q$ *queries to* $\mathbf{VRO}$*.*

The proof of Theorem 6 is deferred to Appendix G for the sake of space.

## References


1. Alagic, G., Carolan, J., Majenz, C., Tokat, S.: The Sponge Is Quantum Indifferentiable. pp. 2591–2630. IEEE Computer Society Press (2025). https://doi.org/10.1109/FOCS63196.2025.00135
2. Amiri-Eliasi, P., Belkheyar, Y., Daemen, J., Ghosh, S., Kuijsters, D., Mehrdad, A., Mella, S., Rasoolzadeh, S., Assche, G.V.: Koala: A Low-Latency Pseudorandom Function. In: Eichlseder, M., Gambs, S. (eds.) Selected Areas in Cryptography - SAC 2024 - 31st International Conference, Montreal, QC, Canada, August 28-30, 2024, Revised Selected Papers, Part II. Lecture Notes in Computer Science, vol. 15517, pp. 239–266. Springer (2024). https://doi.org/10.1007/978-3-031-82841-6_10, https://doi.org/10.1007/978-3-031-82841-6_10
3. Andreeva, E., Bogdanov, A., Dodis, Y., Mennink, B., Steinberger, J.P.: On the Indifferentiability of Key-Alternating Ciphers. In: Canetti, R., Garay, J.A. (eds.) CRYPTO 2013, Part I. LNCS, vol. 8042, pp. 531–550 (Aug 2013). https://doi.org/10.1007/978-3-642-40041-4_29

4. Andreeva, E., Daemen, J., Mennink, B., Van Assche, G.: Security of Keyed Sponge Constructions Using a Modular Proof Approach. In: Leander, G. (ed.) FSE 2015. LNCS, vol. 9054, pp. 364–384. Springer, Heidelberg (Mar 2015). https://doi.org/10.1007/978-3-662-48116-5_18
5. Andreeva, E., Mennink, B., Preneel, B.: Security Reductions of the Second Round SHA-3 Candidates. In: Burmester, M., Tsudik, G., Magliveras, S.S., Ilic, I. (eds.) ISC 2010. LNCS, vol. 6531, pp. 39–53 (Oct 2011). https://doi.org/10.1007/978-3-642-18178-8_5
6. Andreeva, E., Mennink, B., Preneel, B.: The parazoa family: generalizing the sponge hash functions. Int. J. Inf. Sec. **11**(3), 149–165 (2012). https://doi.org/10.1007/S10207-012-0157-6, https://doi.org/10.1007/s10207-012-0157-6
7. Andreeva, E., Stam, M.: The Symbiosis between Collision and Preimage Resistance. In: Chen, L. (ed.) 13th IMA International Conference on Cryptography and Coding. LNCS, vol. 7089, pp. 152–171 (Dec 2011). https://doi.org/10.1007/978-3-642-25516-8_10
8. Aumasson, J.P.: Too Much Crypto. Cryptology ePrint Archive, Report 2019/1492 (2019), https://eprint.iacr.org/2019/1492
9. Bacuieti, N., Daemen, J., Hoffert, S., Van Assche, G., Van Keer, R.: Jammin' on the Deck. pp. 555–584. LNCS (2022). https://doi.org/10.1007/978-3-031-22966-4_19
10. Bellare, M., Jaeger, J., Len, J.: Better Than Advertised: Improved Collision-Resistance Guarantees for MD-Based Hash Functions. In: Thuraisingham, B.M., Evans, D., Malkin, T., Xu, D. (eds.) ACM CCS 2017. pp. 891–906. ACM Press (Oct / Nov 2017). https://doi.org/10.1145/3133956.3134087
11. Ben-David, S., Chor, B., Goldreich, O., Luby, M.: On the Theory of Average Case Complexity. In: 21st ACM STOC. pp. 204–216. ACM Press (May 1989). https://doi.org/10.1145/73007.73027
12. Bertoni, G., Daemen, J., Peeters, M., Van Assche, G.: Sponge Functions. In: ECRYPT hash workshop. vol. 2007 (2007)
13. Bertoni, G., Daemen, J., Peeters, M., Van Assche, G.: On the Indifferentiability of the Sponge Construction. In: Smart, N.P. (ed.) EUROCRYPT 2008. LNCS, vol. 4965, pp. 181–197 (Apr 2008). https://doi.org/10.1007/978-3-540-78967-3_11
14. Bertoni, G., Daemen, J., Peeters, M., Van Assche, G.: Sponge-Based Pseudo-Random Number Generators. In: Mangard, S., Standaert, F.X. (eds.) CHES 2010. LNCS, vol. 6225, pp. 33–47 (Aug 2010). https://doi.org/10.1007/978-3-642-15031-9_3
15. Bertoni, G., Daemen, J., Peeters, M., Van Assche, G.: Duplexing the Sponge: Single-Pass Authenticated Encryption and Other Applications. In: Miri, A., Vaudenay, S. (eds.) SAC 2011. LNCS, vol. 7118, pp. 320–337 (Aug 2012). https://doi.org/10.1007/978-3-642-28496-0_19
16. Bertoni, G., Daemen, J., Peeters, M., Van Assche, G., Van Keer, R., Viguier, B.: KangarooTwelve: Fast Hashing Based on Keccak-p. In: Preneel, B., Vercauteren, F. (eds.) ACNS 18. LNCS, vol. 10892, pp. 400–418 (Jul 2018). https://doi.org/10.1007/978-3-319-93387-0_21
17. Bhattacharjee, A., Chakraborti, A., Datta, N., Mancillas-López, C., Nandi, M.: sf ISAP+: sf ISAP with Fast Authentication. In: Isobe, T., Sarkar, S. (eds.) Progress in Cryptology - INDOCRYPT 2022 - 23rd International Conference on Cryptology in India, Kolkata, India, December 11-14, 2022, Proceedings. Lecture Notes in Computer Science, vol. 13774, pp. 195–219. Springer (2022). https://doi.org/10.1007/978-3-031-22912-1_9, https://doi.org/10.1007/978-3-031-22912-1_9
18. Black, J.: The Ideal-Cipher Model, Revisited: An Uninstantiable Blockcipher-Based Hash Function. In: Robshaw, M.J.B. (ed.) FSE 2006. LNCS, vol. 4047, pp. 328–340. Springer, Heidelberg (Mar 2006). https://doi.org/10.1007/11799313_21
19. Black, J., Rogaway, P., Shrimpton, T., Stam, M.: An Analysis of the Blockcipher-Based Hash Functions from PGV. Journal of Cryptology **23**(4), 519–545 (Oct 2010). https://doi.org/10.1007/s00145-010-9071-0
20. Canetti, R., Chen, Y., Holmgren, J., Lombardi, A., Rothblum, G.N., Rothblum, R.D., Wichs, D.: Fiat-Shamir: from practice to theory. In: Charikar, M., Cohen, E. (eds.) 51st ACM STOC. pp. 1082–1090. ACM Press (Jun 2019). https://doi.org/10.1145/3313276.3316380
21. Canetti, R., Chen, Y., Reyzin, L.: On the Correlation Intractability of Obfuscated Pseudorandom Functions. In: Kushilevitz, E., Malkin, T. (eds.) TCC 2016-A, Part I. LNCS, vol. 9562, pp. 389–415 (Jan 2016). https://doi.org/10.1007/978-3-662-49096-9_17
22. Canetti, R., Chen, Y., Reyzin, L., Rothblum, R.D.: Fiat-Shamir and Correlation Intractability from Strong KDM-Secure Encryption. In: Nielsen, J.B., Rijmen, V. (eds.) EUROCRYPT 2018, Part I. LNCS, vol. 10820, pp. 91–122 (Apr / May 2018). https://doi.org/10.1007/978-3-319-78381-9_4
23. Canetti, R., Goldreich, O., Halevi, S.: The Random Oracle Methodology, Revisited. J. ACM **51**(4), 557–594 (2004). https://doi.org/10.1145/1008731.1008734, https://doi.org/10.1145/1008731.1008734
24. Carolan, J.: Compressed Permutation Oracles. Cryptology ePrint Archive, Paper 2025/1734 (2025), https://eprint.iacr.org/2025/1734
25. Carolan, J., Poremba, A.: Quantum One-Wayness of the Single-Round Sponge with Invertible Permutations. pp. 218–252. LNCS (2024). https://doi.org/10.1007/978-3-031-68391-6_7
26. Choudhuri, A.R., Garg, S., Jain, A., Jin, Z., Zhang, J.: Correlation Intractability and SNARGs from Sub-exponential DDH. pp. 635–668. LNCS (2023). https://doi.org/10.1007/978-3-031-38551-3_20

27. Cogliati, B., Seurin, Y.: On the Provable Security of the Iterated Even-Mansour Cipher Against Related-Key and Chosen-Key Attacks. In: Oswald, E., Fischlin, M. (eds.) EUROCRYPT 2015, Part I. LNCS, vol. 9056, pp. 584–613 (Apr 2015). https://doi.org/10.1007/978-3-662-46800-5_23
28. Cojocaru, A., Hhan, M., Liu, Q., Yamakawa, T., Yun, A.: Quantum Lifting for Invertible Permutations and Ideal Ciphers. In: Kalai, Y.T., Kamara, S.F. (eds.) Advances in Cryptology - CRYPTO 2025 - 45th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 17-21, 2025, Proceedings, Part II. Lecture Notes in Computer Science, vol. 16001, pp. 481–512. Springer (2025). https://doi.org/10.1007/978-3-032-01878-6_16, https://doi.org/10.1007/978-3-032-01878-6_16
29. Cooper, D.A., Apon, D.C., Dang, Q.H., Davidson, M.S., Dworkin, M.J., Miller, C.A., et al.: Recommendation for Stateful Hash-Based Signature Schemes. NIST Special Publication **800**(208), 800–208 (2020)
30. Coretti, S., Dodis, Y., Karthikeyan, H., Tessaro, S.: Seedless Fruit Is the Sweetest: Random Number Generation, Revisited. In: Boldyreva, A., Micciancio, D. (eds.) CRYPTO 2019, Part I. LNCS, vol. 11692, pp. 205–234 (Aug 2019). https://doi.org/10.1007/978-3-030-26948-7_8
31. Coron, J.S., Dodis, Y., Malinaud, C., Puniya, P.: Merkle-Damgård Revisited: How to Construct a Hash Function. In: Shoup, V. (ed.) CRYPTO 2005. LNCS, vol. 3621, pp. 430–448 (Aug 2005). https://doi.org/10.1007/11535218_26
32. Coron, J.S., Holenstein, T., Künzler, R., Patarin, J., Seurin, Y., Tessaro, S.: How to Build an Ideal Cipher: The Indifferentiability of the Feistel Construction. Journal of Cryptology **29**(1), 61–114 (Jan 2016). https://doi.org/10.1007/s00145-014-9189-6
33. DAEMEN, J.: Deck-function-based Cryptography. Symmetric Cryptography, Volume 2: Cryptanalysis and Future Directions p. 227 (2023)
34. Daemen, J., Mennink, B., Van Assche, G.: Full-State Keyed Duplex with Built-In Multi-user Support. In: Takagi, T., Peyrin, T. (eds.) ASIACRYPT 2017, Part II. LNCS, vol. 10625, pp. 606–637 (Dec 2017). https://doi.org/10.1007/978-3-319-70697-9_21
35. Dodis, Y., Ristenpart, T., Shrimpton, T.: Salvaging Merkle-Damgård for Practical Applications. In: Joux, A. (ed.) EUROCRYPT 2009. LNCS, vol. 5479, pp. 371–388 (Apr 2009). https://doi.org/10.1007/978-3-642-01001-9_22
36. Dworkin, M.J.: SP 800-38D. Recommendation for block cipher modes of operation: Galois/Counter Mode (GCM) and GMAC. National Institute of Standards & Technology (2007)
37. Dworkin, M.J., et al.: SHA-3 standard: Permutation-based hash and extendable-output functions (2015)
38. Eaton, E.: Leighton-Micali Hash-Based Signatures in the Quantum Random-Oracle Model. Cryptology ePrint Archive, Report 2017/607 (2017), https://eprint.iacr.org/2017/607
39. Fluhrer, S.: Further Analysis of a Proposed Hash-Based Signature Standard. Cryptology ePrint Archive, Report 2017/553 (2017), https://eprint.iacr.org/2017/553
40. Foekens, R.: Security of the Sponge Construction with a Random Transformation (2023)
41. Fuchs, J., Rotella, Y., Daemen, J.: On the Security of Keyed Hashing Based on Public Permutations. pp. 607–627. LNCS (2023). https://doi.org/10.1007/978-3-031-38548-3_20
42. Geest, D.V., Bashiri, K., Fluhrer, S., Gazdag, S., Kousidis, S.: RFC 9802: Use of the HSS and XMSS Hash-Based Signature Algorithms in Internet X.509 Public Key Infrastructure (2025)
43. Grover, L.K.: A Fast Quantum Mechanical Algorithm for Database Search. In: 28th ACM STOC. pp. 212–219. ACM Press (May 1996). https://doi.org/10.1145/237814.237866
44. Gueron, S., Lindell, Y.: Better Bounds for Block Cipher Modes of Operation via Nonce-Based Key Derivation. In: Thuraisingham, B.M., Evans, D., Malkin, T., Xu, D. (eds.) ACM CCS 2017. pp. 1019–1036. ACM Press (Oct / Nov 2017). https://doi.org/10.1145/3133956.3133992
45. Guido, B., Joan, D., Michaël, P., Gilles, V.: Cryptographic Sponge Functions. 2011-STMicroelectronics NXP Semiconductors, Version 0.1 January 14 (2011)
46. Gunsing, A., Daemen, J., Mennink, B.: Deck-Based Wide Block Cipher Modes and an Exposition of the Blinded Keyed Hashing Model. IACR Trans. Symm. Cryptol. **2019**(4), 1–22 (2019). https://doi.org/10.13154/tosc.v2019.i4.1-22
47. Guo, J., Peyrin, T., Poschmann, A.: The PHOTON Family of Lightweight Hash Functions. In: Rogaway, P. (ed.) CRYPTO 2011. LNCS, vol. 6841, pp. 222–239 (Aug 2011). https://doi.org/10.1007/978-3-642-22792-9_13
48. Hirose, S., Park, J.H., Yun, A.: A Simple Variant of the Merkle-Damgård Scheme with a Permutation. Journal of Cryptology **25**(2), 271–309 (Apr 2012). https://doi.org/10.1007/s00145-010-9095-5
49. Hoffman, P., Schneier, B.: RFC 4270: Attacks on cryptographic hashes in internet protocols. Network Working Group (2005)
50. Housley, R.: RFC 9708: Use of the HSS/LMS Hash-Based Signature Algorithm in the Cryptographic Message Syntax (CMS) (2025)
51. Hu, K., Niu, Z., Wang, M.: Round-Based Approximation of (Higher-Order) Differential-Linear Correlation – A Geometric Approach Perspective. In: Advances in Cryptology - EUROCRYPTO 2026, to appear. Lecture Notes in Computer Science, vol. 16004
52. Kalai, Y.T., Rothblum, G.N., Rothblum, R.D.: From Obfuscation to the Security of Fiat-Shamir for Proofs. In: Katz, J., Shacham, H. (eds.) CRYPTO 2017, Part II. LNCS, vol. 10402, pp. 224–251 (Aug 2017). https://doi.org/10.1007/978-3-319-63715-0_8

53. Katz, J.: Analysis of a proposed hash-based signature standard. In: International conference on research in security standardisation. pp. 261–273. Springer (2016)
54. Kelsey, J., Schneier, B.: Second Preimages on n-Bit Hash Functions for Much Less than 2n Work. In: Cramer, R. (ed.) EUROCRYPT 2005. LNCS, vol. 3494, pp. 474–490 (May 2005). https://doi.org/10.1007/11426639_28
55. Lamport, L.: Constructing digital signatures from a one way function (1979)
56. Leander, N.G., Mennink, B., Naya-Plasencia, M., Sasaki, Y., Lambooij, E.: Symmetric Cryptography (Dagstuhl Seminar 22141). Dagstuhl Reports **12**(4), 1–12 (2022). https://doi.org/10.4230/DAGREP.12.4.1, https://doi.org/10.4230/DagRep.12.4.1
57. Lefevre, C.: Indifferentiability of the Sponge Construction with a Restricted Number of Message Blocks. IACR Trans. Symm. Cryptol. **2023**(1), 224–243 (2023). https://doi.org/10.46586/tosc.v2023.i1.224-243
58. Lefevre, C., Belkheyar, Y., Daemen, J.: Kirby: A Robust Permutation-Based PRF Construction. Cryptology ePrint Archive, Report 2023/1520 (2023), https://eprint.iacr.org/2023/1520
59. Lefevre, C., Mennink, B.: Tight Preimage Resistance of the Sponge Construction. pp. 185–204. LNCS (2022). https://doi.org/10.1007/978-3-031-15985-5_7
60. Lefevre, C., Mennink, B.: Permutation-Based Hashing Beyond the Birthday Bound. IACR Trans. Symm. Cryptol. **2024**(1), 71–113 (2024). https://doi.org/10.46586/tosc.v2024.i1.71-113
61. Leighton, F.T., Micali, S.: Large provably fast and secure digital signature schemes based on secure hash functions (Jul 11 1995), uS Patent 5,432,852
62. Lombardi, A., Vaikuntanathan, V.: Correlation-Intractable Hash Functions via Shift-Hiding. pp. 102:1–102:16. LIPIcs (2022). https://doi.org/10.4230/LIPIcs.ITCS.2022.102
63. Majenz, C., Malavolta, G., Walter, M.: Permutation Superposition Oracles for Quantum Query Lower Bounds. pp. 1508–1519. ACM Press (2025). https://doi.org/10.1145/3717823.3718266
64. McGrew, D., Curcio, M., Fluhrer, S.: RFC 8554: Leighton-Micali Hash-Based Signatures (2019)
65. Mennink, B.: Keying Merkle-Damgård at the Suffix. IACR Trans. Symmetric Cryptol. **2025**(1), 70–96 (2025). https://doi.org/10.46586/TOSC.V2025.I1.70-96, https://doi.org/10.46586/tosc.v2025.i1.70-96
66. Mennink, B., Neves, S.: Encrypted Davies-Meyer and Its Dual: Towards Optimal Security Using Mirror Theory. In: Katz, J., Shacham, H. (eds.) CRYPTO 2017, Part III. LNCS, vol. 10403, pp. 556–583 (Aug 2017). https://doi.org/10.1007/978-3-319-63697-9_19
67. Mennink, B., Preneel, B.: On the Impact of Known-Key Attacks on Hash Functions. In: Iwata, T., Cheon, J.H. (eds.) ASIACRYPT 2015, Part II. LNCS, vol. 9453, pp. 59–84 (Nov / Dec 2015). https://doi.org/10.1007/978-3-662-48800-3_3
68. Merkle, R.C.: A Certified Digital Signature. In: Brassard, G. (ed.) CRYPTO'89. LNCS, vol. 435, pp. 218–238 (Aug 1990). https://doi.org/10.1007/0-387-34805-0_21
69. Merkle, R.C.: Secrecy, authentication, and public key systems. Stanford university (1979)
70. Mogul, J., Hoff, A.V.: RFC 3230: Instance Digests in HTTP (2002)
71. Naya-Plasencia, M.: Symmetric Cryptography for Long Term Security. EUROCRYPT 2022 Invited talk (2022), https://eurocrypt.iacr.org/2022/slides/Eurocrypt22-NayaPlasencia.pdf.
72. Neven, G., Smart, N.P., Warinschi, B.: Hash function requirements for Schnorr signatures. J. Math. Cryptol. **3**(1), 69–87 (2009). https://doi.org/10.1515/JMC.2009.004, https://doi.org/10.1515/JMC.2009.004
73. NICCS: Announcement on Launching the Next-generation Commercial Cryptographic Algorithms Program (2025), https://www.niccs.org.cn/niccs/Notice/pc/content/content_1937427069632253952.html.
74. NIST: Lightweight Cryptography (February 2019). Cryptology ePrint Archive, Paper 2022/1033 (2019), https://csrc.nist.gov/Projects/Lightweight-Cryptography.
75. Niu, Z., Hu, K., Sun, S., Zhang, Z., Wang, M.: Speeding Up Preimage and Key-Recovery Attacks with Highly Biased Differential-Linear Approximations. pp. 73–104. LNCS (2024). https://doi.org/10.1007/978-3-031-68385-5_3
76. Peikert, C., Shiehian, S.: Noninteractive Zero Knowledge for NP from (Plain) Learning with Errors. In: Boldyreva, A., Micciancio, D. (eds.) CRYPTO 2019, Part I. LNCS, vol. 11692, pp. 89–114 (Aug 2019). https://doi.org/10.1007/978-3-030-26948-7_4
77. Peyrin, T., Seurin, Y.: Counter-in-Tweak: Authenticated Encryption Modes for Tweakable Block Ciphers. Version 20160524:153228. Cryptology ePrint Archive, Report 2015/1049 (2015), https://eprint.iacr.org/archive/2015/1049/20151029:213132.
78. Pub, F.: Secure Hash Standard (SHS). Fips pub **180**(4), 2012 (2012)
79. Rogaway, P.: Formalizing Human Ignorance. In: Nguyen, P.Q. (ed.) Progressin Cryptology - VIETCRYPT 2006, First International Conference on Cryptology in Vietnam, Hanoi, Vietnam, September 25-28, 2006, Revised Selected Papers. Lecture Notes in Computer Science, vol. 4341, pp. 211–228. Springer (2006). https://doi.org/10.1007/11958239_14, https://doi.org/10.1007/11958239_14
80. Rogaway, P., Shrimpton, T.: Cryptographic Hash-Function Basics: Definitions, Implications, and Separations for Preimage Resistance, Second-Preimage Resistance, and Collision Resistance. In: Roy, B.K., Meier, W. (eds.) FSE 2004. LNCS, vol. 3017, pp. 371–388. Springer, Heidelberg (Feb 2004). https://doi.org/10.1007/978-3-540-25937-4_24
81. Rubin, A.: RFC 1805: Location-Independent Data/Software Integrity Protocol (1995)

82. Soni, P., Tessaro, S.: Public-Seed Pseudorandom Permutations. In: Coron, J., Nielsen, J.B. (eds.) EUROCRYPT 2017, Part II. LNCS, vol. 10211, pp. 412–441 (Apr / May 2017). https://doi.org/10.1007/978-3-319-56614-6_14
83. Soni, P., Tessaro, S.: Naor-Reingold Goes Public: The Complexity of Known-Key Security. In: Nielsen, J.B., Rijmen, V. (eds.) EUROCRYPT 2018, Part III. LNCS, vol. 10822, pp. 653–684 (Apr / May 2018). https://doi.org/10.1007/978-3-319-78372-7_21
84. Sönmez Turan, M., McKay, K., Chang, D., Kang, J., Kelsey, J.: Ascon-Based Lightweight Cryptography Standards for Constrained Devices: Authenticated Encryption, Hash, and Extendable Output Functions. Tech. rep., National Institute of Standards and Technology (2024)
85. Stelvio Cimato, Joan Daemen, S.M.G.V.A.: Permutation-based Crypto 2025. May 4, 2025 - Madrid, Spain. Co-located with EUROCRYPT 2025 (2025), https://permutationbasedcrypto.org/2025/.
86. Stinson, D.R.: Some observations on the theory of cryptographic hash functions. Cryptology ePrint Archive, Report 2001/020 (2001), https://eprint.iacr.org/2001/020
87. Sun, S., Li, S., Zhang, Z., Lefevre, C., Mennink, B., Qin, Z., Feng, D.: Permutation-Based Hashing With Stronger (Second) Preimage Resistance. Cryptology ePrint Archive, Paper 2025/963, version 20250526:220706 (2025), https://eprint.iacr.org/archive/2025/963/20250526:220706
88. Sun, S., Li, S., Zhang, Z., Lefevre, C., Mennink, B., Qin, Z., Feng, D.: Permutation-Based Hashing with Stronger (Second) Preimage Resistance-Application to Hash-Based Signature Schemes. Cryptology ePrint Archive (2025)
89. Viguier, B., Wong, D., Van Assche, G., Dang, Q., Daemen, J.: RFC 9861 – KangarooTwelve and TurboSHAKE (2025)
90. Wuille, P., Nick, J., Ruffing, T.: Schnorr Signatures for secp256k1. BIP340 (2020), https://github.com/bitcoin/bips/blob/master/bip-0340.mediawiki, retrieved 26/09/2025
91. Yasuda, K.: How to Fill Up Merkle-Damgård Hash Functions. In: Pieprzyk, J. (ed.) ASIACRYPT 2008. LNCS, vol. 5350, pp. 272–289 (Dec 2008). https://doi.org/10.1007/978-3-540-89255-7_17

# A Applying Human Ignorance and CI to Standard Sponge

We also applied the Human Ignorance approach to the standard sponge: however, the underlying evasive relations (something similar to the "multi-block CICO problem" [45, Sect. 8.2.5]) are quite complicated and less informative. We sketch the results in this appendix. In this respect, feed-forward creates somewhat independence between different permutation invocations and significantly simplifies the relations.

Concretely, we consider the standard sponge construction in Fig. 1 with absorbing capacity $c_1$ and rate $r_1$, and focus on the case of squeezing a single block output $Z = Z[1]$, $|Z| = r_2$ for simplicity. In addition, we focus on the plain model, i.e., the "oracle-less" setting, since we do not derive random permutation bounds.

**Preimage security (as per the definition in Sect. 2.1).** For this, for every $Z \in \{0,1\}^{r_2}$ we consider a $\ell_{\max}$-ary relation $R_{\text{pre-sp}}(Z)$:

$$\begin{aligned}&\big((X_1,...,X_{\ell_{\max}}),(Y_1,...,Y_{\ell_{\max}})\big) \in R_{\text{pre-sp}}(Z)\\ &\textbf{iff } \exists \ell \in \{1,...,\ell_{\max}\} : \big(\mathsf{right}_{c_1}(X_1) = IV\big) \wedge \big(\mathsf{left}_{r_2}(Y_\ell) = Z\big)\\ &\quad \wedge \big(\mathsf{right}_{c_1}(Y_i) = \mathsf{right}_{c_1}(X_{i+1}) \text{ for all } i = 1,...,\ell-1\big).\end{aligned} \tag{49}$$

It can be seen that breaking sponge preimage security is related to breaking $R_{\text{pre-sp}}(Z)$-ci, i.e., $\mathbf{Adv}^{Z\text{-pre}}_{\mathsf{Sponge}^{\mathsf{P},\mathsf{pd}}}(\mathscr{A}) \le \mathbf{Adv}^{R_{\text{pre-sp}}(Z)\text{-ci}}_{\mathsf{P}}(\mathscr{B})$.

Breaking $R_{\text{pre-sp}}(Z)$-ci of $\mathsf{P}$ quite resembles solving the "multi-block CICO problem" discussed in [45, Sect. 8.2.5]. However, $R_{\text{pre-sp}}(Z)$ is complicated, and **breaking $R_{\text{pre-sp}}(Z)$-ci does not appear conceptually simpler than finding a sponge preimage** for $Z$. This means the reduction in this section is less informative and convincing than the reductions for $\mathsf{Sponge}\text{-}\mathsf{F}^{\mathsf{P},\mathsf{pd}}$ (Sect. 4.2). In addition, since $R_{\text{pre-sp}}(Z)$ is $\ell_{\max}$-ary, the simpler approaches adopted in Sect. 4.2 to deriving quantum bounds cannot yield meaning bounds. However, due to the invertibility of evaluations in $\mathsf{Sponge}^{\mathsf{P},\mathsf{pd}}(M)$ (given an entire $b$-bit intermediate state), we are unable to find relations that are both simpler than $R_{\text{pre-sp}}(Z)$ and sufficient for the reduction.

**Second preimage security (as defined in Sect. 2.1).** For every $Y \in \{0,1\}^b$ we consider a $\ell_{\max}$-ary relation $R_{\text{spr-sp}}(X,Y)$:

$$\begin{aligned}&\big((X'_1,...,X'_{\ell_{\max}}),(Y'_1,...,Y'_{\ell_{\max}})\big) \in R_{\text{spr-sp}}(Y)\\ &\textbf{iff } \exists \ell \in \{1,...,\ell_{\max}\} : \big(\mathsf{right}_{c_1}(X'_1) = IV\big)\\ &\quad \wedge \Big(Y'_\ell \ne Y \wedge \big(\mathsf{left}_{r_2}(Y'_\ell) = \mathsf{left}_{r_2}(Y) \vee \mathsf{right}_{c_1}(Y'_\ell) = \mathsf{right}_{c_1}(Y)\big)\Big)\\ &\quad \wedge \big(\mathsf{right}_{c_1}(Y'_i) = \mathsf{right}_{c_1}(X'_{i+1}) \text{ for all } i = 1,...,\ell-1\big).\end{aligned} \tag{50}$$

Built on this, we introduce the following family of relations indexed by $Y$:

$$\mathcal{R}_{\text{spr-sp}} := \big\{R_{\text{spr-sp}}(Y),\ Y \in \{0,1\}^b\big\}. \tag{51}$$

It can be seen that a challenge message $M$ with $\mathsf{pd}(M)$ consisting of $\ell$ blocks can be compiled into $\ell$ $b$-bit challenges $Y_1, ..., Y_\ell$ (to wit, they are the $b$-bit intermediate values of $\mathsf{Sponge}^{\mathsf{P},\mathsf{pd}}(M)$), whereas an adversary $\mathscr{A}(M)$ outputting $M' \neq M$ such that $\mathsf{Sponge}^{\mathsf{P},\mathsf{pd}}(M') = \mathsf{Sponge}^{\mathsf{P},\mathsf{pd}}(M)$ can be transformed into a $(\mathcal{R}_{\text{spr-sp}}(Y_1), ..., \mathcal{R}_{\text{spr-sp}}(Y_\ell))$-mci adversary $\mathscr{B}(Y_1, ..., Y_{\ell_2})$ giving rise to $\big((X'_1, ..., X'_{\ell_{\max}}), (P(X'_1), ..., P(X'_{\ell_{\max}}))\big) \in R_{\text{spr-sp}}(Y_i)$ for some $i \in \{1, ..., \ell_2\}$. Concretely, $\mathscr{B}(Y_1, ..., Y_{\ell_2})$ recovers $M$ from its inputs, and runs $\mathscr{A}(M) \Rightarrow M'$ to have $M'$. $\mathscr{B}$ then computes $X'_1, Y'_1 = \mathsf{P}(X'_1), ..., X'_\ell, Y'_\ell = \mathsf{P}(X'_\ell)$ by evaluating $\mathsf{Sponge}^{\mathsf{P},\mathsf{pd}}(M')$ and sets the remaining $X'_{\ell+1}, ..., X'_{\ell_{\max}}$ to arbitrary values, and it suffices to output these $X'_1, ..., X'_{\ell_{\max}}$. By $\mathsf{Sponge}^{\mathsf{P},\mathsf{pd}}(M') = \mathsf{Sponge}^{\mathsf{P},\mathsf{pd}}(M)$, it can be seen that $\big((X'_1, ..., X'_{\ell_{\max}}), (P(X'_1), ..., P(X'_{\ell_{\max}}))\big) \in R_{\text{spr-sp}}(Y_i)$ indeed holds for some index $i \leq \ell_2$.

The relation $R_{\text{spr-sp}}(Y)$ is an extension of the relation $R_{\text{pre-sp}}(Z)$ for preimage security, and is also related to the "multi-block CICO problem" of [45, Sect. 8.2.5]. Again, $R_{\text{spr-sp}}(Y)$ is complicated and less informative.

**Collision security.** We consider a $2\ell_{\max}$-ary relation $R_{\text{coll-sp}}$:

$$\begin{aligned}
&\big((X_1, ..., X_{\ell_{\max}}, X'_1, ..., X'_{\ell_{\max}}), (Y_1, ..., Y_{\ell_{\max}}, Y'_1, ..., Y'_{\ell_{\max}})\big) \in R_{\text{coll-sp}} \\
&\textbf{iff } \exists \ell_1, \ell_2 \in \{1, ..., \ell_{\max}\} : \big(X_1 \neq X'_1 \wedge \mathsf{right}_{c_1}(X_1) = \mathsf{right}_{c_1}(X'_1)\big) \\
&\quad \wedge \big(Y_{\ell_1} \neq Y'_{\ell_2} \wedge \mathsf{left}_{r_2}(Y_{\ell_1}) = \mathsf{left}_{r_2}(Y'_{\ell_2}) \vee \mathsf{right}_{c_1}(Y_{\ell_1}) = \mathsf{right}_{c_1}(Y'_{\ell_2})\big) \\
&\quad \wedge \big(\mathsf{right}_{c_1}(Y_i) = \mathsf{right}_{c_1}(X_{i+1}) \text{ for all } i = 1, ..., \ell_1 - 1\big) \\
&\quad \wedge \big(\mathsf{right}_{c_1}(Y'_i) = \mathsf{right}_{c_1}(X'_{i+1}) \text{ for all } i = 1, ..., \ell_2 - 1\big).
\end{aligned} \tag{52}$$

It can be seen that a collision adversary $\mathscr{A}$ outputting $M \neq M'$ with $\mathsf{Sponge}^{\mathsf{P},\mathsf{pd}}(M) = \mathsf{Sponge}^{\mathsf{P},\mathsf{pd}}(M')$ can be transformed into an adversary $\mathscr{B}$ against $R_{\text{coll-sp}}$-ci. Again, $R_{\text{coll-sp}}$ is rather complicated and informativeless.

# B Previous Works on Constructing CI Keyed Functions

Previous works mostly focused on CI of (hash) functions rather than permutations, and we thus focus on (hash) functions in this appendix. Consider a *keyless* (hash) function $\mathsf{h} : \{0,1\}^* \mapsto \{0,1\}^h$. Similarly to the discussion in Sect. 2.2, in the typical non-uniform adversarial model, as long as the relation $R$ is non-trivial, i.e., $(X_1, ..., X_m)$ such that $\big((X_1, ..., X_m), (\mathsf{h}(X_1), ..., \mathsf{h}(X_m))\big) \in R$, no keyless function $\mathsf{h}$ can have "negligible" $R$-ci advantage

$$\mathbf{Adv}^{R\text{-ci}}_{\mathsf{h}}(\mathscr{A}) := \Pr\big[\mathscr{A}(R) \Rightarrow (X_1, ..., X_m), \big((X_1, ..., X_m), (\mathsf{h}(X_1), ..., \mathsf{h}(X_m))\big) \in R\big].$$

To this end, previous works [23,21] focused on *keyed (hash) functions* $\mathsf{h} : \mathcal{HK} \times \{0,1\}^* \mapsto \{0,1\}^h$ (they used the terminology *function ensemble* of theoretical community, but we stick with *keyed function* that is more commonly used in symmetric cryptography). In addition, we use the concrete security paradigm without explicit security parameters.

**Correlation intractability for keyed functions.** Given a keyed (hash) function $\mathsf{h} : \mathcal{HK} \times \{0,1\}^* \mapsto \{0,1\}^h$, a relation $R$ and an adversary $\mathscr{A}$, the advantage of $\mathscr{A}$ against the $R$-ci of $\mathsf{h}$ is defined as

$$\begin{aligned}
\mathbf{Adv}^{R\text{-ci}}_{\mathsf{h}}(\mathscr{A}) := \Pr\big[&hk \xleftarrow{\$} \mathcal{HK} : \mathscr{A}(hk) \Rightarrow (X_1, ..., X_m), \\
&\big((X_1, \mathsf{P}(X_1)), ..., (X_m, \mathsf{P}(X_m))\big) \in R\big].
\end{aligned}$$

Asymptotically, $\mathsf{h}$ is $R$-ci w.r.t. $R$, if $\mathbf{Adv}^{R\text{-ci}}_{\mathsf{h}}(\mathscr{A})$ is a negligible function of the security parameter for all efficient $\mathscr{A}$.

**Positive results on CI keyed functions.** Canetti et al. [21] constructed a correlation intractable keyed function that withstands all *unary relations* (which was called *binary relations* in [21]) recognizable in a-priori bounded polynomial complexity. Namely, for any fixed polynomial $p$, they construct a keyed function as follows: for any evasive (see below) relation $R$ computable in time $p$, given a random key $hk \xleftarrow{\$} \mathcal{HK}$, it is hard to find $X$ such that $(X, f(hk, X)) \in R$. Kalai et al. [52] and then Canetti et al. [22] used strong forms of key-dependent message secure encryption schemes to construct correlation intractable keyed functions that withstand all unary relations, and then Canetti et al. [20] constructed correlation intractable keyed functions that withstand all efficiently sampleable (or computable) unary relations. Subsequent works [76,62,26] managed to enlarge the involved relation family and weaken the underlying assumptions.

Going beyond unary relations, Lombardi and Vaikuntanathan [62] managed to enlarge the involved relation family and weaken the underlying assumptions.

Restricting to relations recognizable in a-priori bounded polynomial complexity is somewhat *necessary* for efficiently achieving CI [23,21]. In this respect, note that the relations $R_{\text{pre}}(Z)$, $R_{\text{spr}}(X, \nu)$ and $R_{\text{coll}}$ used in Sect. 4.2 are indeed recognizable by circuits with size $Cb$ for a small constant $C$. Combined with the positive results regarding CI

w.r.t. bounded-complexity relations, this provides some justifications for collision and (second) preimage security of $\mathsf{Sponge}\text{-}\mathsf{F}^{\mathsf{P},\mathsf{pd}}$, $\mathsf{Sponge}\text{-}\mathsf{P}^{\mathsf{P},\mathsf{pd}}$ and the underlying permutation-based truncated Davies-Meyer constructions from a theoretical point of view (namely, a permutation satisfying $R_{\mathrm{pre}}(Z)$-ci, $R_{\mathrm{spr}}(X,\nu)$-ci and $R_{\mathrm{coll}}$-ci may indeed exists in theory), which complements justifications from cryptanalytic results.

# C Brief Preliminary on Quantum World

A quantum system $Q$ is defined over a finite set $B$ of classical states. The pure states of $Q$ form a complex Hilbert space of dimension $|B|$, where the vectors/states assign a complex weight to each element in $B$. We will denote pure states using the key notation $\phi\rangle$. Given quantum systems $Q_0, Q_1$ over $B_0, B_1$ respectively, the product system is $Q = Q_0 \times Q_1$ over states $B = B_0 \times B_1 = \{(b_0, b_1) : b_0 \in B_0, b_1 \in B_1\}$.

A pure state $|\phi\rangle$ is manipulated by performing a unitary transformation $U$ to the state $|\phi\rangle$. A pure state $|\phi\rangle$ can also be measured; the measurement outputs the value $x$ with probability $|\langle x \mid \phi\rangle|^2$; afterward, the state "collapses" to the state $|x\rangle$. A quantum computer will be able to perform a fixed, finite set $G$ of unitary transformations, which we will call gates. For concreteness, we will use so-called Hadamard, phase, CNOT and $\pi/8$ gates. Each gate or measurement costs unit time to apply. An efficient quantum algorithm will be able to make a polynomial-length sequence of operations, where each operation is either a gate from $G$ or a measurement.

Any efficiently computable classical function can also be computed efficiently on a quantum computer, though care is needed to make the transformation unitary. Concretely, if $f$ is computable by a polynomial-sized circuit, then there is a efficiently computable unitary $U_f$ on the quantum system $Q = Q_{\mathrm{in}} \otimes Q_{\mathrm{out}} \otimes Q_{\mathrm{work}}$ with the property that: $U_f|x, y, 0\rangle = |x, y+f(x), 0\rangle$. Here, $Q_{\mathrm{work}}$ is an ancillary quantum system that is used as workspace, and is erased after the computation.

If a quantum algorithm has quantum oracle access to a function $f$, the oracle applies the unitary $U_f$ as defined above.

# D Deferred Relevant Hash Security Definitions

Andreeva and Stam [7, Sect. 3] relabeled the preimage resistance notion w.r.t. randomly chosen message $M$ as *domain-oriented preimage resistance* (the definition can be found in [80, Definition 1]), and the preimage resistance notion w.r.t. randomly chosen image $Z$ as *range-oriented preimage resistance* (the definition can be found in [19]). To capture applications in which the challenge messages/images are not uniformly distributed, Andreeva and Stam incorporated distributions into the definitions, yielding *distribution-oriented preimage resistance*. Formally, Andreeva and Stam's definition *for keyed hash* functions is recalled as follows.

**Definition 2 ([7], Definition 2).** *Consider a keyed hash function $\mathsf{h} : \mathcal{HK} \times \{0,1\}^* \to \mathcal{Z}$ with key space $\mathcal{HK}$. Let $\mathbf{q}$ be a distribution over $\mathcal{Z}$ and $\mathbf{p}$ a distribution over $\{0,1\}^*$. The domain-oriented and range-oriented preimage finding advantage of adversary $\mathscr{A}$ are defined as*

$$\mathbf{Adv}^{\mathbf{p}\text{-dpre}}_{\mathsf{h}}(\mathscr{A}) := \Pr\big[hk \xleftarrow{\$} \mathcal{HK}, M \xleftarrow{\$} \mathbf{p}, Z \leftarrow \mathsf{h}(hk, M), \mathscr{A}(hk, Z) \Rightarrow M' : M' \neq M \wedge \mathsf{h}(hk, M') = Z\big], \tag{53}$$

$$\mathbf{Adv}^{\mathrm{rpre}\text{-}\mathbf{q}}_{\mathsf{h}}(\mathscr{A}) := \Pr\big[hk \xleftarrow{\$} \mathcal{HK}, Z \xleftarrow{\$} \mathbf{q}, \mathscr{A}(hk, Z) \Rightarrow M : \mathsf{h}(hk, M) = Z\big]. \tag{54}$$

In [7, page 158], it was explicitly required that the distribution $\mathbf{q}$ over $\mathcal{Z}$ used in $\mathbf{Adv}^{\mathrm{rpre}\text{-}\mathbf{q}}_{\mathsf{h}}(\mathscr{A})$ is *independent of the hash key $hk$*. No such explicit restriction was mentioned regarding $\mathbf{p}$ in $\mathbf{Adv}^{\mathbf{p}\text{-dpre}}_{\mathsf{h}}(\mathscr{A})$, but the definition in Eq. (53) appears to insist on this key-independence.

We remark that domain-oriented preimage deviates from second preimage resistance: in the former setting, the adversary $\mathscr{A}$ merely gets the challenge image $Z$ as input, while in the latter setting $\mathscr{A}$ gets the full challenge message $M$ as input (which means $\mathscr{A}$ can compute all intermediate values). Finally, while Definition 2 focused on keyed hash, it is straightforward to adapt it to keyless hash, and it can be seen that the advantages could be sufficiently small even for non-uniform adversaries as long as the distributions $\mathbf{p}$ and $\mathbf{q}$ are "sufficiently uniform".

# E Deferred Proof for $\mathsf{Sponge}\text{-}\mathsf{F}^{\mathsf{P},\mathsf{pd}}$

## E.1 Proof of Lemma 1

Take $(\mathsf{A}, \mathsf{B}) = (Z\text{-pre}, R_{\mathrm{pre}}(Z)\text{-ci})$ for example. Regardless of $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ is classical or quantum, the two claims are equivalent: (i) The $b$-bit output $X$ of $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ breaks $Z$-pre security of $\mathsf{TrDM}^{\mathsf{P}}_h$; (ii) The $b$-bit output $X$ of $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ has $\mathsf{right}_h\big(X \oplus \mathsf{P}(X)\big) = Z$, i.e., $(X, \mathsf{P}(X)) \in R_{\mathrm{pre}}(Z)$.

**Procedure** $\textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}(M)$
$(\overline{M}[1],\ldots,\overline{M}[\ell]) \xleftarrow{r} \mathsf{pd}(M)$, $\overline{Y}_0 \leftarrow 0^r \| IV$
**for** $i = 1, 2, ..., \ell$ **do**
  **if** $i \neq \ell$ **then**
    $X_i \leftarrow \overline{Y}_{i-1} \oplus (\overline{M}[i] \| 0^c)$
  **else** // $i = \ell$
    $X_i \leftarrow \overline{Y}_{i-1} \oplus (\overline{M}[i] \| \theta)$
  $S \leftarrow \mathsf{right}_c(X_i)$, $Y_i \leftarrow \mathsf{P}(X_i)$, $\overline{Y}_i \leftarrow Y_i \oplus (0^r \| S)$
**return** $(X_1, Y_1, ..., X_\ell, Y_\ell)$

Fig. 5: Procedure $\textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}$ that compute and return the intermediate values of $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}(M, h)$.

**Algorithm** $\mathscr{B}_{\mathrm{coll}}^{\mathsf{P},\mathsf{P}^{-1}}$
$(M_1, M_2) \leftarrow \mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$
$(X_1, Y_1, ..., X_{\ell_1}, Y_{\ell_1}) \leftarrow \textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}(M_1)$
$(X'_1, Y'_1, ..., X'_{\ell_2}, Y'_{\ell_2}) \leftarrow \textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}(M_2)$
**return** $\textsc{FindCollision}\big((X_1, Y_1, ..., X_{\ell_1}, Y_{\ell_1}), (X'_1, Y'_1, ..., X'_{\ell_2}, Y'_{\ell_2})\big)$

**Algorithm** $\mathscr{B}_{\mathrm{pre}}^{\mathsf{P},\mathsf{P}^{-1}}$
$(M_1, M_2) \leftarrow \mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$
$(X_1, Y_1, ..., X_{\ell_1}, Y_{\ell_1}) \leftarrow \textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}(M_1)$
$(X'_1, Y'_1, ..., X'_{\ell_2}, Y'_{\ell_2}) \leftarrow \textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}(M_2)$
*coll* $\leftarrow \textsc{FindCollision}\big((X_1, Y_1, ..., X_{\ell_1}, Y_{\ell_1}), (X'_1, Y'_1, ..., X'_{\ell_2}, Y'_{\ell_2})\big)$
**if** *coll* $\neq \perp$ **then**
  **return** $\perp$
// Now it holds $X_{\ell_1 - \min\{\ell_1,\ell_2\}+1} = X'_{\ell_2 - \min\{\ell_1,\ell_2\}+1} = \star \| IV$, and it cannot be $\ell_1 = \ell_2$.
**if** $\ell_1 > \ell_2$ **then**
  **return** $X_{\ell_1 - \ell_2}$
**else** // $\ell_1 < \ell_2$
  **return** $X'_{\ell_2 - \ell_1}$

**Procedure** $\textsc{FindCollision}\big((X_1, Y_1, ..., X_{\ell_1}, Y_{\ell_1}), (X'_1, Y'_1, ..., X'_{\ell_2}, Y'_{\ell_2})\big)$
$i \leftarrow \min\{\ell_1, \ell_2\}$
**for** $j = 0, 1, ..., i-1$ **do**
  **if** $X_{\ell_1 - j} \neq X'_{\ell_2 - j}$ **then**
    **return** $(X_{\ell_1 - j}, X'_{\ell_2 - j})$
**return** $\perp$

Fig. 6: Adversaries $\mathscr{B}_{\mathrm{coll}}^{\mathsf{P},\mathsf{P}^{-1}}$ and $\mathscr{B}_{\mathrm{pre}}^{\mathsf{P},\mathsf{P}^{-1}}$ for the proof of Theorem 2. They use a procedure $\textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}$ defined in Fig. 5.

### E.2 Proof of Lemma 2

Consider $\mathsf{X}$ = classical first. The two adversaries $\mathscr{B}_{\mathrm{coll}}$ and $\mathscr{B}_{\mathrm{pre}}$ are given in Fig. 6. The claims on complexities are clear by construction. To prove Eq. (20), consider the view of adversary $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ when run by either $\mathscr{B}_{\mathrm{coll}}^{\mathsf{P},\mathsf{P}^{-1}}$ or $\mathscr{B}_{\mathrm{pre}}^{\mathsf{P},\mathsf{P}^{-1}}$. If $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ successfully finds a collision $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}(M_1, h) = \mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}(M_2, h)$, then the intermediate values $(X_1, Y_1, ..., X_{\ell_1}, Y_{\ell_1}) \leftarrow \textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}(M_1)$ and $(X'_1, Y'_1, ..., X'_{\ell_2}, Y'_{\ell_2}) \leftarrow \textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}(M_2)$ consist of two cases.

*Case 1: There exists* $j \in \{0, ..., \min\{\ell_1, \ell_2\} - 1\}$ *such that* $X_{\ell_1 - j} \neq X'_{\ell_2 - j}$. Let $j$ be the smallest integer such that $X_{\ell_1 - j} \neq X'_{\ell_2 - j}$. If $j = 0$ then $\mathsf{TrDM}_h^{\mathsf{P}}(X_{\ell_1}) = \mathsf{TrDM}_h^{\mathsf{P}}(X'_{\ell_2}) = Z$. Otherwise, we have $X_{\ell_1 - j + 1} = X'_{\ell_2 - j + 1}$, i.e., $\mathsf{TrDM}_h^{\mathsf{P}}(X_{\ell_1 - j}) = \mathsf{TrDM}_h^{\mathsf{P}}(X'_{\ell_2 - j})$. By construction, the call to $\textsc{FindCollision}$ will find this $j$ in the for loop at line 22 and return $(X_{\ell_1 - j}, X'_{\ell_2 - j})$ to $\mathscr{B}_{\mathrm{coll}}^{\mathsf{P},\mathsf{P}^{-1}}$, meaning that $\mathbf{Adv}^{\mathrm{coll}}_{\mathsf{TrDM}_h^{\mathsf{P}}}(\mathscr{B}_{\mathrm{coll}}) \geq \mathbf{Adv}^{\mathrm{coll}}_{\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}}(\mathscr{A})$. On the other hand, in this case, $\mathscr{B}_{\mathrm{pre}}^{\mathsf{P},\mathsf{P}^{-1}}$ outputs $\perp$ in the for loop at line 13, meaning that $\mathbf{Adv}^{IV\text{-}\mathrm{pre}}_{\mathsf{TrDM}_c^{\mathsf{P}}}(\mathscr{B}_{\mathrm{pre}}) = 0$. Therefore, $\mathbf{Adv}^{\mathrm{coll}}_{\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}}(\mathscr{A}) \leq \mathbf{Adv}^{\mathrm{coll}}_{\mathsf{TrDM}_h^{\mathsf{P}}}(\mathscr{B}_{\mathrm{coll}}) + \mathbf{Adv}^{IV\text{-}\mathrm{pre}}_{\mathsf{TrDM}_c^{\mathsf{P}}}(\mathscr{B}_{\mathrm{pre}})$ holds in this case.

*Case 2:* $X_{\ell_1-j} = X'_{\ell_2-j}$ *for all* $j \in \{0,...,\min\{\ell_1,\ell_2\}-1\}$. This means $\ell_1 \neq \ell_2$: otherwise, it necessarily holds $M_1 = M_2$ and they do not constitute a valid collision pair. Wlog assume $\ell_1 > \ell_2$: the case of $\ell_1 < \ell_2$ is similar by symmetry. Then, $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}(M_1,h) = \mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}(M_2,h)$ does not give rise to any collision of $\mathsf{TrDM}_h^{\mathsf{P}}$, and $\mathscr{B}_{\mathrm{coll}}^{\mathsf{P},\mathsf{P}^{-1}}$ returns $\perp$ (which is returned by its call to FindCollision at line 25). On the other hand, since $X_{\ell_1-\ell_2+1} = X'_1$ and $\mathsf{right}_c(X'_1) = IV$, it holds $\mathsf{TrDM}_c^{\mathsf{P}}(X_{\ell_1-\ell_2}) = IV$ and the value $X_{\ell_1-\ell_2}$ returned by $\mathscr{B}_{\mathrm{pre}}^{\mathsf{P},\mathsf{P}^{-1}}$ at line 16 has $\mathsf{TrDM}_c^{\mathsf{P}}(X_{\ell_1-\ell_2}) = IV$. Note that the arguments also hold in the case where $\ell_2 = 1$. Therefore, $\mathbf{Adv}^{\mathrm{coll}}_{\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}}(\mathscr{A}) \leq \mathbf{Adv}^{\mathrm{coll}}_{\mathsf{TrDM}_h^{\mathsf{P}}}(\mathscr{B}_{\mathrm{coll}}) + \mathbf{Adv}^{IV\text{-pre}}_{\mathsf{TrDM}_c^{\mathsf{P}}}(\mathscr{B}_{\mathrm{pre}})$ also holds in this case.

The above established Eq. (20). The second claim Eq. (21) then directly follows by Lemma 1.

The above arguments also hold in the quantum setting $\mathsf{X}$ = quantum (for simplicity, the classical queries made via $\textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}$ are also counted into quantum query complexity): note that we never utilize techniques that are difficult to handle in the quantum setting, including rewinding, recording quantum queries, etc.

### E.3 Preimage Security of $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$

**Lemma 4 (Preimage).** *Let* $(\mathsf{P},\mathsf{P}^{-1})$ *be a permutation oracle (not necessarily random) and its inverse,* $\mathsf{X} \in \{$*classical*, *quantum*$\}$*, and let* $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ *be a* $(q,t,s,\ell_{\max})$*-bounded* $\mathsf{X}$ *preimage adversary against* $\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}$*. Then, we can construct a* $(q+\ell_{\max}, t+O(\ell_{\max}), s+O(\ell_{\max}))$*-bounded* $\mathsf{X}$ *preimage adversary* $\mathscr{B}_{\mathrm{pre}}^{\mathsf{P},\mathsf{P}^{-1}}$ *against* $\mathsf{TrDM}_h^{\mathsf{P}}$*, such that the following holds for any* $Z \in \{0,1\}^h$*:*

$$\mathbf{Adv}^{Z\text{-pre}}_{\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}}(\mathscr{A}) \leq \mathbf{Adv}^{Z\text{-pre}}_{\mathsf{TrDM}_h^{\mathsf{P}}}(\mathscr{B}_{\mathrm{pre}}) = \mathbf{Adv}^{R_{\mathrm{pre}}(Z)\text{-ci}}_{\mathsf{P}}(\mathscr{B}_{\mathrm{pre}}). \tag{55}$$

*Proof.* $\mathscr{B}_{\mathrm{pre}}$ runs $\mathscr{A}(Z)$ to have $M$ with $\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}(M) = Z$, computes underlying calls $\mathsf{P}(X_1) = Y_1, ..., \mathsf{P}(X_\ell) = Y_\ell$, and outputs $X_\ell$, which has $\mathsf{TrDM}_h^{\mathsf{P}}(X_\ell) = Z$ as long as $\mathscr{A}(Z)$ succeeds. This proves $\mathbf{Adv}^{Z\text{-pre}}_{\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}}(\mathscr{A}) \leq \mathbf{Adv}^{Z\text{-pre}}_{\mathsf{TrDM}_h^{\mathsf{P}}}(\mathscr{B}_{\mathrm{pre}})$, which equals $\mathbf{Adv}^{R_{\mathrm{pre}}(Z)\text{-ci}}_{\mathsf{P}}(\mathscr{B}_{\mathrm{pre}})$ by Lemma 1. Similarly to Theorem 2, the arguments also hold in the quantum setting. □

### E.4 Proof of Lemma 3

Consider $\mathsf{X}$ = classical first.

$\textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}$ is given in Fig. 5. For $(X_1,...,X_\ell) \leftarrow \textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}(M)$, the distinctness of $X_1,...,X_\ell$ directly follows by $M$ being free-of-inner-collision.

The adversaries $\mathscr{B}_{\mathrm{mspr}^*}^{\mathsf{P},\mathsf{P}^{-1}}$ and $\mathscr{B}_{\mathrm{spr}}^{\mathsf{P},\mathsf{P}^{-1}}$ are given in Fig. 7. Their complexities are clear by construction (note that reconstructing $M$ from $X_1,...,X_\ell$ consumes $\ell \leq \ell_{\max}$ queries, while $\textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}$ consumes at most $\ell_{\max}$ queries).

To prove Eq. (22), consider the view of adversary $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ when run by one of the adversaries. If $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}(M)$ successfully finds $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}(M',h) = \mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}(M,h)$, then the values $(X'_1,Y'_1,...,X'_{\ell_2},Y'_{\ell_2}) \leftarrow \textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}(M_2)$ consist of three cases.

*Case 1: there exists* $j \in \{1,...,\min\{\ell,\ell_2\}-1\}$ *such that* $X_{\ell-j} \neq X'_{\ell_2-j}$. Let $j$ be the smallest integer such that $X_{\ell-j} \neq X'_{\ell_2-j}$. When $j=0$, it holds $\mathsf{right}_h(X_\ell \oplus Y_\ell) = \mathsf{Sponge\text{-}P}^{\mathsf{P},\mathsf{pd}}(M,h) = \mathsf{Sponge\text{-}P}^{\mathsf{P},\mathsf{pd}}(M',h) = \mathsf{right}_h(X'_{\ell_2} \oplus Y'_{\ell_2})$, and $X'_{\ell_2}$ constitutes a second preimage for $\mathsf{TrDM}_h^{\mathsf{P}}$ for challenge $X_\ell$. When $j \geq 1$, it holds $\mathsf{right}_c(X_{\ell-j} \oplus Y_{\ell-j}) = \mathsf{right}_c(X_{\ell-j+1}) = \mathsf{right}_c(X'_{\ell_2-j}) = \mathsf{right}_c(X'_{\ell_2-j} \oplus Y'_{\ell_2-j})$, and $X'_{\ell_2-j}$ constitutes a second preimage for $\mathsf{TrDM}_c^{\mathsf{P}}$ for the challenge input $X_{\ell-j}$. By construction, $\mathscr{B}_{\mathrm{mspr}^*}^{\mathsf{P},\mathsf{P}^{-1}}$ outputs this $X'_{\ell_2-j}$, while $\mathscr{B}_{\mathrm{pre}}^{\mathsf{P},\mathsf{P}^{-1}}$ outputs $\perp$. Therefore, Eq. (22) also holds in this case.

*Case 2:* $X_{\ell-j} = X'_{\ell_2-j}$ *for all* $j \in \{0,...,\min\{\ell,\ell_2\}-1\}$. This means $\ell \neq \ell_2$: otherwise, $M' = M$ is not a valid second preimage. Meanwhile, it cannot be $\ell > \ell_2$: otherwise, it holds $\mathsf{right}_c(X_{\ell-\ell_2+1}) = \mathsf{right}_c(X'_1) = IV$ and contradicts the assumption that $M$ is free-of-IV-preimage. By this, it holds $\ell_2 > \ell$, and the value $X'_{\ell_2-\ell}$ returned by $\mathscr{B}_{\mathrm{pre}}^{\mathsf{P},\mathsf{P}^{-1}}$ has $\mathsf{TrDM}_c^{\mathsf{P}}(X'_{\ell_2-\ell}) = \mathsf{right}_c(X'_{\ell_2-\ell+1}) = IV$ and provides a preimage for $\mathsf{TrDM}_c^{\mathsf{P}}$ for the image $IV$. The arguments also hold for $\ell = 1$. Eq. (22) thus also holds.

In all, Eq. (22) holds in any case. This plus Lemma 1 yield Eq. (23). Similarly to Lemma 2, the arguments also hold when $\mathsf{X}$ = quantum.

**Algorithm** $\mathscr{B}^{\mathsf{P},\mathsf{P}^{-1}}_{\mathrm{mspr}^*}(X_1,...,X_\ell)$
Reconstruct $M$ from $X_1,...,X_\ell$
$M' \leftarrow \mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}(M)$
$(X'_1,Y'_1,...,X'_{\ell_2},Y'_{\ell_2}) \leftarrow \textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}(M')$
$coll \leftarrow \textsc{FindCollision}^{\mathsf{P},\mathsf{pd}}\big((X_1,Y_1,...,X_\ell,Y_\ell),(X'_1,Y'_1,...,X'_{\ell_2},Y'_{\ell_2})\big)$
**if** $coll = (X_{\ell-j}, X'_{\ell_2-j})$ **then**
  **return** $X'_{\ell_2-j}$
**return** $\bot$

**Algorithm** $\mathscr{B}^{\mathsf{P},\mathsf{P}^{-1}}_{\mathrm{pre}}(X_1,...,X_\ell)$
Reconstruct $M$ from $X_1,...,X_\ell$
$M' \leftarrow \mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}(M)$
$(X'_1,Y'_1,...,X'_{\ell_2},Y'_{\ell_2}) \leftarrow \textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}(M')$
$coll \leftarrow \textsc{FindCollision}^{\mathsf{P},\mathsf{pd}}\big((X_1,Y_1,...,X_\ell,Y_\ell),(X'_1,Y'_1,...,X'_{\ell_2},Y'_{\ell_2})\big)$
**if** $coll \neq \bot$ **or** $\ell > \ell_2$ **then**
  **return** $\bot$
**return** $X'_{\ell_2-\ell}$

Fig. 7: Adversaries $\mathscr{B}^{\mathsf{P},\mathsf{P}^{-1}}_{\mathrm{mspr}^*}$ and $\mathscr{B}^{\mathsf{P},\mathsf{P}^{-1}}_{\mathrm{spr}}$ for the proof of Theorem 3. They use the procedure $\textsc{FindCollision}^{\mathsf{P},\mathsf{pd}}$ defined in Fig. 6.

### E.5 Proof of Theorem 3

Let $\mathscr{S}$ be classical algorithm that samples $\mathbf{d}$.

The classical adversary $\mathscr{B}_{\mathrm{mspr}^*}$ runs $\mathscr{S}$ to have a challenge $M$, evaluates $\mathsf{Sponge}\text{-}\mathsf{F}^{\mathsf{P},\mathsf{pd}}_h(M)$ to have the underlying calls $\mathsf{P}(X_1)=Y_1,...,\mathsf{P}(X_\ell)=Y_\ell$, and outputs $(X_i,X_j)$ that has $\mathsf{TrDM}^{\mathsf{P}}_h(X_i)=\mathsf{TrDM}^{\mathsf{P}}_h(X_j)$. Clearly, $\mathscr{B}_{\mathrm{coll}}$ consumes $t_D+O(\ell_{\max})$ time and $s_D+O(\ell_{\max})$ space, and succeeds if and only if the sampled $M$ is *not* free-of-inner-collision (denote this event by $\mathsf{MInCol}$).

The $\mathsf{X}$ adversary $\mathscr{B}_{\mathrm{pre}}$ runs $\mathscr{S}$ to have a challenge $M$, runs $\mathscr{A}(M)\Rightarrow M'$ to have a second preimage $M'$ of $M$, and evaluates $\mathsf{Sponge}\text{-}\mathsf{F}^{\mathsf{P},\mathsf{pd}}_h(M)$ and $\mathsf{Sponge}\text{-}\mathsf{F}^{\mathsf{P},\mathsf{pd}}_h(M')$ to have underlying calls $\mathsf{P}(X_1)=Y_1,...,\mathsf{P}(X_\ell)=Y_\ell;\mathsf{P}(X'_1)=Y'_1,...,\mathsf{P}(X'_{\ell_2})=Y'_{\ell_2}$, and outputs the value $X_i$ with $\mathsf{TrDM}^{\mathsf{P}}_c(X_i)=IV$ or $X'_i$ with $\mathsf{TrDM}^{\mathsf{P}}_c(X'_i)=IV$. $\mathscr{B}_{\mathrm{pre}}$ consumes $t_D+O(\ell_{\max})$ time and $s_D+O(\ell_{\max})$ space, and succeeds if and only if the sampled $M$ is *not* free-of-IV-preimage *or* the second preimage $M'$ returned by $\mathscr{A}$ indicates $(\mathsf{TrDM}^{\mathsf{P}}_c)^{-1}(IV)$ (denote this event by $\mathsf{IVPre}$).

The $(t_D+O(\ell_{\max}),s_D+O(\ell_{\max}))$-bounded sampler $\mathscr{S}_2$ for the distribution $\mathbf{d}_2$: runs $\mathscr{S}\Rightarrow M$ to sample $M$ according to $\mathbf{d}$, evaluates $\mathsf{Sponge}\text{-}\mathsf{F}^{\mathsf{P},\mathsf{pd}}_h(M)$ to have underlying calls $\mathsf{P}(X_1)=Y_1,...,\mathsf{P}(X_\ell)=Y_\ell$ and outputs $X_1,...,X_\ell$.

Finally, the adversary $\mathscr{B}_{\mathrm{mspr}^*}$ proceeds the same as the adversary $\mathscr{B}^{\mathsf{P},\mathsf{P}^{-1}}_{\mathrm{mspr}^*}$ described for Lemma 3. By Lemma 3, $\mathscr{B}_{\mathrm{mspr}^*}$ wins as long as the challenges $X_1,...,X_\ell$ are generated using a free-of-inner-collision message $M$. Therefore,

$$\begin{aligned}\mathbf{Adv}^{\mathbf{d}\text{-spr}}_{\mathsf{Sponge}\text{-}\mathsf{F}^{\mathsf{P},\mathsf{pd}}_h}(\mathscr{A}) &\le \Pr\big[\mathsf{MInCol}\vee\mathsf{IVPre}\big]+\Pr\big[\mathscr{B}_{\mathrm{mspr}^*}\text{ succeeds }\mid\neg\mathsf{MInCol}\wedge\neg\mathsf{IVPre}\big]\\ &\le \mathbf{Adv}^{\mathrm{coll}}_{\mathsf{TrDM}^{\mathsf{P}}_h}(\mathscr{B}_{\mathrm{coll}})+\mathbf{Adv}^{IV\text{-pre}}_{\mathsf{TrDM}^{\mathsf{P}}_c}(\mathscr{B}_{\mathrm{pre}})+\mathbf{Adv}^{\mathbf{d}_2\text{-mspr}^*}_{\mathsf{TrDM}^{\mathsf{P}}_c}(\mathscr{B}_{\mathrm{mspr}^*}).\end{aligned}$$

This establishes Eq. (27). Eq. (28) then follows by Lemma 1.

Again, the arguments also hold when $\mathsf{X}=$ quantum—but the collision adversary $\mathscr{B}_{\mathrm{coll}}$ and the sampler $\mathscr{S}'$ remain classical.

## F Proof of Theorem 5

Fluhrer [39, Theorem 1] provided a lms-mfpspr security proof for SHA-256 in the random compression function model. However, as will be shown in App. I.2, his proof is *flawed*. Therefore, while we could follow his general idea (which is actually sound), we have to give a more involved dedicated analysis in the random permutation model.

Consider an experiment $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge}\text{-}\mathsf{F}^{\mathbf{\Pi},\mathsf{apd}}_h,\mathscr{A},\mu,\Phi}$ in which there are $Q$ queries to $\mathbf{\Pi}$ and $\mathbf{\Pi}^{-1}$ in total. We maintain a table $T$ for these $\mathbf{\Pi}$ queries and responses. Concretely, the table $T$ map entries $X\in\{0,1\}^b$ to $Y\in\{0,1\}^b$, and is empty at the beginning. Every time an $\mathbf{\Pi}$-query $\mathbf{\Pi}(X)\to Y$ or $\mathbf{\Pi}^{-1}(Y)\to X$ occurs in $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge}\text{-}\mathsf{F}^{\mathbf{\Pi},\mathsf{apd}}_h,\mathscr{A},\mu,\Phi}$, a *table entry* $T(X)=Y$ is defined.

Recall from Theorem 5 that every prefix $\mathsf{prefix}_{i,j}$ is of the form $I^{\mathsf{idx}}\|\star$ for some $\mathsf{idx} \in \{1, ..., u\}$, where $u$ is an integer determined by LMS, and every $I^{\mathsf{idx}}$ has $|I^{\mathsf{idx}}| = 2r$. As will be demonstrated in App. I.3, $I^1, ..., I^u$ are actually the $2r$-bit "user identifiers" of the $u$ involved LMS signature users. A part of the table $T$ records the $\mathbf{\Pi}$-queries for processing these $2r$-bit "user identifiers". Formally, let $T_I$ be the partial table such that $T_I(X) = Y$ is defined if and only if:

- $\mathsf{right}_c(X) = IV$ and there exists $\mathsf{idx} \in \{1, ..., u\}$ such that $\mathsf{left}_r(X) = \mathsf{left}_r(I^{\mathsf{idx}})$ for the $\mathsf{idx}$-th user's identifier $I^{\mathsf{idx}}$; or
- There exists $\mathsf{idx}$ such that $\mathsf{left}_r(X) = \mathsf{right}_r(I^{\mathsf{idx}})$ for the $\mathsf{idx}$-th user's identifier $I^{\mathsf{idx}}$, and $\mathsf{right}_c(X) = IV \oplus \mathsf{right}_c(Y')$ for the table entry $T_I(\mathsf{left}_r(I^{\mathsf{idx}})\|IV) = Y'$.

It clearly holds $|T_I| \leq 2u$.

Below we first define and analyze several bad events that may occur during an experiment $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}},\mathscr{A},\mu,\Phi}$. Experiments during which no bad event occurs are called *good*. We then analyze properties of such good experiments, with the help of which we are able to derive an upper bound on the success probability of $\mathscr{A}$.

## F.1 Bad Events and Probabilities

We define several bad events that may occur during experiment $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}},\mathscr{A},\mu,\Phi}$.

- $\mathsf{HitIV}$: There exists a table entry $T(X) = Y$ such that $\mathsf{right}_c(X \oplus Y) = IV$ or $\mathsf{right}_c(X \oplus Y) = IV \oplus \theta$;
- $\mathsf{FixP}$: There exists a table entry $T(X) = Y$ that is defined after a forward query $\mathbf{\Pi}(X) \to Y$, such that $\mathsf{right}_c(Y) = 0^c$ or $\mathsf{right}_c(Y) = \theta$;
- $\mathsf{BadLW}$: There exist two distinct entries $T(X_1) = Y_1$ and $T(X_2) = Y_2$ such that: (i) $\mathsf{right}_c(X_1 \oplus Y_1) \in \{\mathsf{right}_c(X_2), \mathsf{right}_c(X_2) \oplus \theta\}$, and (ii) $\mathsf{right}_c(Y_2) \in \{0^c, \theta\}$, and (iii) $T(X_2) = Y_2$ is defined after a backward query $\mathbf{\Pi}^{-1}(Y_2) \to X_2$, which happens after $T(X_1) = Y_1$ is defined.
- $\mathsf{3coll}$: There exist three distinct entries $T(X_1) = Y_1, T(X_2) = Y_2, T(X_3) = Y_3$ and three bits $\beta_1, \beta_2, \beta_3 \in \{0, 1\}$ such that $\mathsf{right}_c(X_1 \oplus Y_1) \oplus \beta_1 \cdot \theta = \mathsf{right}_c(X_2 \oplus Y_2) \oplus \beta_2 \cdot \theta = \mathsf{right}_c(X_3 \oplus Y_3) \oplus \beta_3 \cdot \theta$;
- $\mathsf{MColPair}$: There exist $C$ distinct pairs $\big(T(X_1) = Y_1, T(X_1') = Y_1'\big), ..., \big(T(X_C) = Y_C, T(X_C') = Y_C'\big)$ such that $\mathsf{right}_c(X_i \oplus Y_i) = \mathsf{right}_c(X_i' \oplus Y_i')$ or $\mathsf{right}_c(X_i \oplus Y_i) = \mathsf{right}_c(X_i' \oplus Y_i') \oplus \theta$ for all $i \in \{1, ..., C\}$;
- $\mathsf{ICol}$: There exist distinct records $T_I(X) = Y, T_I(X') = Y'$ in $T_I$ with $\mathsf{right}_c(X \oplus Y) = \mathsf{right}_c(X' \oplus Y')$ or $\mathsf{right}_c(X \oplus Y) = \mathsf{right}_c(X' \oplus Y') \oplus \theta$;
- $\mathsf{hCol}$: There exist $h$ distinct pairs of distinct entries $\big(T(X_1) = Y_1, T(X_1') = Y_1'\big), ...,$ $\big(T(X_h) = Y_h, T(X_h') = Y_h'\big)$ and $h$ bits $\beta_1, ..., \beta_h \in \{0, 1\}$ such that

$$\mathsf{right}_c(X_1 \oplus Y_1 \oplus X_1' \oplus Y_1') \oplus \beta_1 \cdot \theta = ... = \mathsf{right}_c(X_h \oplus Y_h \oplus X_h' \oplus Y_h') \oplus \beta_h \cdot \theta.$$

If the total number of $\mathbf{\Pi}$-queries appeared during $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}},\mathscr{A},\mu,\Phi}$ is $Q$, then it is easy to see (note that $h = c$)

$$\Pr\big[\mathsf{HitIV}\big] \leq 2 \cdot \frac{2^r Q}{2^b - Q} \leq \frac{4Q}{2^h}, \tag{56}$$

$$\Pr\big[\mathsf{FixP}\big] \leq 2 \cdot \frac{2^r Q}{2^b - Q} = \frac{4Q}{2^h}, \tag{57}$$

$$\Pr\big[\mathsf{BadLW}\big] \leq Q \times 2^{r+1} \times \frac{2^r}{2^b - Q} \leq \frac{2^{r+2} Q}{2^h}, \tag{58}$$

$$\Pr\big[\mathsf{3coll}\big] \leq \binom{Q}{3} \cdot \Big(\frac{2 \cdot 2^r}{2^b - Q}\Big)^2 \leq \frac{4Q^3}{2^{2h}}, \tag{59}$$

$$\Pr\big[\mathsf{MColPair}\big] \leq \Big(\binom{Q}{2} \cdot \frac{4}{2^c}\Big)^C \cdot \frac{1}{C!} \leq \Big(\frac{2Q^2}{2^h}\Big)^C \cdot \frac{1}{C!}. \tag{60}$$

Regarding $\mathsf{ICol}$, it is easy to see $|\mathcal{Q}_I| \leq 2u$, which immediately implies

$$\Pr\big[\mathsf{ICol}\big] \leq \binom{2u}{2} \cdot \frac{2 \cdot 2^r}{2^b - Q} \leq \frac{8u^2}{2^h}. \tag{61}$$

Finally, for any pair of distinct entries $\big(T(X_i) = Y_i, T(X_i') = Y_i'\big), ..., \big(T(X_j) = Y_j, T(X_j') = Y_j'\big)$, it can be seen the probability to have $\mathsf{right}_c(X_i \oplus Y_i \oplus X_i' \oplus Y_i') = \mathsf{right}_c(X_j \oplus Y_j \oplus X_j' \oplus Y_j')$ or $\mathsf{right}_c(X_i \oplus Y_i \oplus X_i' \oplus Y_i') = \mathsf{right}_c(X_j \oplus Y_j \oplus X_j' \oplus Y_j') \oplus \theta$ is at most $4/2^c$. Since the number of such pairs is at most $Q^2$, when $4Q^2 \leq 2^h$ it holds

$$\Pr\big[\mathsf{hCol}\big] \leq \binom{Q^2}{h} \cdot \Big(\frac{2}{2^h}\Big)^{h-1} \leq \Big(\frac{8Q^2}{2^h}\Big)^h \cdot \frac{1}{h!} \leq \frac{8Q^2}{h! 2^h}. \tag{62}$$

For simplicity, let $\mathsf{Bad} := \mathsf{HitIV} \vee \mathsf{FixP} \vee \mathsf{BadLW} \vee \mathsf{3coll} \vee \mathsf{MColPair} \vee \mathsf{ICol} \vee \mathsf{hCol}$. A union bound yields

$$\begin{aligned}\Pr[\mathsf{Bad}] &\leq \frac{4Q}{2^h} + \frac{4Q}{2^h} + \frac{2^{r+2}Q}{2^h} + \frac{4Q^3}{2^{2h}} + \left(\frac{2Q^2}{2^h}\right)^C \cdot \frac{1}{C!} + \frac{8u^2}{2^h} + \frac{8Q^2}{h!2^h} \\ &\leq \frac{8Q + 2^{r+2}Q + 8u^2}{2^h} + \frac{4Q^3}{2^{2h}} + \left(\frac{2Q^2}{2^h}\right)^C \cdot \frac{1}{C!} + \frac{8Q^2}{h!2^h}. \end{aligned} \tag{63}$$

If $\mathsf{Bad}$ occurs then we say the experiment $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}},\mathscr{A},\mu,\Phi}$ is *bad*; otherwise, we say it is *good*.

### F.2 Analysis of Good Experiments

**Transcript induced graph.** For $S \in \{0,1\}^c$, define $\mathbb{V}_S$ as the set $\mathbb{V}_S := \{S, S{\oplus}\theta\}$. With this notation, we define a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ over entries in $T$:

(i) The vertex set of $\mathcal{G}$ is $\mathcal{V} = \{\mathbb{V}_S : S \in \{0,1\}^c\}$;
(ii) For every table entry $T(X) = Y$, $\mathcal{E}$ contains an edge from $\mathbb{V}_{\mathsf{right}_c(X)}$ to $\mathbb{V}_{\mathsf{right}_c(X \oplus Y)}$.

Adjacent edges in $\mathcal{G}$ would form *paths*. Concretely, a sequence of $\ell$ table entries

$$T(X_1) = Y_1,\ T(X_2) = Y_2, ...,\ T(X_\ell) = Y_\ell$$

satisfying $\mathsf{right}_c(X_{i+1}) = \mathsf{right}_c(X_i{\oplus}Y_i)$ or $\mathsf{right}_c(X_{i+1}) = \mathsf{right}_c(X_i{\oplus}Y_i){\oplus}\theta$ for all $i = 1, ..., \ell - 1$ form a *length-$\ell$ path* from the vertex $\mathbb{V}_{\mathsf{right}_c(X_1)}$ to the vertex $\mathbb{V}_S$, $S = \mathsf{right}_c(X_\ell{\oplus}Y_\ell)$. We denote this path by $\left(\mathbb{V}_{\mathsf{right}_c(X_1)} \xrightarrow{\overline{M}} \mathbb{V}_S\right)$, where $\overline{M} = \overline{M}[1]\|...\|\overline{M}[\ell]$, $\overline{M}[1] = \mathsf{left}_r(X_1)$, $\overline{M}[i+1] = \mathsf{left}_r(Y_i){\oplus}\mathsf{left}_r(X_{i+1})$ for $i = 1, ..., \ell - 1$.

A special form of path is of particular interest. In detail, for $\ell \geq 2$, a sequence of $\ell$ table entries (or $\ell$ edges)

$$T(X_1) = Y_1,\ T(X_2) = Y_2, ...,\ T(X_\ell) = Y_\ell$$

satisfying:

(i) $\mathsf{right}_c(X_1) = IV$, and
(ii) $\mathsf{right}_c(X_{i+1}) = \mathsf{right}_c(X_i{\oplus}Y_i)$ for all $i = 1, ..., \ell - 2$, and
(iii) $\mathsf{right}_c(X_\ell) = \mathsf{right}_c(X_{\ell-1}{\oplus}Y_{\ell-1}){\oplus}\theta$

form a *length-$\ell$ hashing path*, with starting point $IV = \mathsf{right}_c(X_1)$ and endpoint $Z = \mathsf{right}_c(X_\ell{\oplus}Y_\ell)$. We denote this path by $\left(\overline{M}, Z, 1\right)$ (note that this is consistent with the notation to-be-used in Sect. G), where $\overline{M} = \overline{M}[1]\|...\|\overline{M}[\ell]$, $\overline{M}[1] = \mathsf{left}_r(X_1)$, $\overline{M}[i+1] = \mathsf{left}_r(Y_i){\oplus}\mathsf{left}_r(X_{i+1})$ for $i = 1, ..., \ell - 1$. Such a hashing path $\left(\overline{M}, Z, 1\right)$ with $\mathsf{unapd}(\overline{M}) = M \neq \perp$ captures the computation of $\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}}(M)$. In addition, it indicates the existence of a length-$\ell$ path $\left(\mathbb{V}_{IV} \xrightarrow{\overline{M}} \mathbb{V}_Z\right)$ in the graph $\mathcal{G}$.

With this terminology, $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}},\mathscr{A},\mu,\Phi}$ generating a challenge digest $Z^{i,j} \leftarrow \mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}}(\mathsf{prefix}_{i,j}\|M^{i,j})$ indicates creating "challenge path" $\left(\mathbb{V}_{IV} \xrightarrow{\mathsf{apd}(\mathsf{prefix}_{i,j}\|M^{i,j})} \mathbb{V}_{Z^{i,j}}\right)$ in $\mathcal{G}$. We introduce two second preimage-related events, i.e., *post-challenge second preimage event* $\mathsf{PostChSpr2}$ and *pre-challenge second preimage event* $\mathsf{PreChSpr2}$:

– $\mathsf{PostChSpr2}$: After a challenge path $\left(\mathbb{V}_{IV} \xrightarrow{\mathsf{apd}(\mathsf{prefix}_{i,j}\|M^{i,j})} \mathbb{V}_{Z^{i,j}}\right)$, $i \in \{1,2,3\}$, $j \in \{1, ..., \gamma_i\}$, is created in **Phase 2** of the experiment $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}},\mathscr{A},\mu,\Phi}$, another query $\mathbf{\Pi}(X) \to Y$ or $\mathbf{\Pi}^{-1}(Y) \to X$ creates a colliding path $\left(\mathbb{V}_{IV} \xrightarrow{\mathsf{apd}(\mathsf{prefix}_{i,j}\|M')} \mathbb{V}_{Z^{i,j}}\right)$ with $M' \neq M^{i,j}$;
– $\mathsf{PreChSpr2}$: When a challenge path $\left(\mathbb{V}_{IV} \xrightarrow{\mathsf{apd}(\mathsf{prefix}_{i,j}\|M^{i,j})} \mathbb{V}_{Z^{i,j}}\right)$, $i \in \{1,2,3\}$, $j \in \{1, ..., \gamma_i\}$, is newly created in **Phase 2** of the experiment $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}},\mathscr{A},\mu,\Phi}$, there has been a corresponding colliding path $\left(\mathbb{V}_{IV} \xrightarrow{\mathsf{apd}(\mathsf{prefix}_{i,j}\|M')} \mathbb{V}_{Z^{i,j}}\right)$ in $\mathcal{G}$, with $M' \neq M^{i,j}$.

By the definition of $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}},\mathscr{A},\mu,\Phi}$, the adversary's success is mostly due to the occurrence of $\mathsf{PostChSpr2}$ or $\mathsf{PreChSpr2}$.

**On user identifier values.** The $u$ aforementioned "user identifiers" $I^1, ..., I^u$ may not be distinct. Assume that the number of distinct values among the $u$ user identifiers $I^1, ..., I^u$ is $\alpha_I$, and denote these values by

$$II_1, ..., II_{\alpha_I}.$$

Since $|II_i| = 2r$, they induce $\alpha_I$ length-2 paths in $\mathcal{G}$. Let $\left(\mathbb{V}_{IV} \xrightarrow{II_1} \mathbb{V}_{S_1^{II}}\right), ..., \left(\mathbb{V}_{IV} \xrightarrow{II_{\alpha_I}} \mathbb{V}_{S_{\alpha_I}^{II}}\right)$ be these $\alpha_I$ paths, then conditioned on $\neg\mathsf{IColl}$, it can be seen that $\mathbb{V}_{S_1^{II}}, ..., \mathbb{V}_{S_{\alpha_I}^{II}}$ are pairwise distinct.

**Lemma 5.** *In a good experiment* $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}},\mathscr{A},\mu,\Phi}$*, for any two distinct user identifier values* $II_i$ *and* $II_j$*, the corresponding paths* $\left(\mathbb{V}_{IV} \xrightarrow{II_i} \mathbb{V}_{S_i^{II}}\right)$ *and* $\left(\mathbb{V}_{IV} \xrightarrow{II_j} \mathbb{V}_{S_j^{II}}\right)$ *necessarily have* $\mathbb{V}_{S_i^{II}} \neq \mathbb{V}_{S_j^{II}}$ *(which implies* $S_i \neq S_j$*). (This is easily seen from* $\neg\mathsf{IColl}$ *and we omit the detailed proof.)*

**On collided paths.** Let $S, Z \in \{0,1\}^c$. The vertex $\mathbb{V}_S$ *can be reached from* $\mathbb{V}_{IV}$ *via the k-th prefix value* $PP_k$, if there exists a path of the form $\left(\mathbb{V}_{IV} \xrightarrow{PP_k\|\star} \mathbb{V}_S\right)$. The vertex $\mathbb{V}_S$ *has a "limited path" to another (not necessarily distinct) vertex* $\mathbb{V}_Z$, if there exists a path of the form $\left(\mathbb{V}_S \xrightarrow{\overline{M}} \mathbb{V}_Z\right)$ with length at most $\ell_{\max} - 1$ (i.e., $|\overline{M}| \leq (\ell_{\max} - 1)r$). With these terminology, we put forward a combinatorial lemma as follows.

**Lemma 6.** *In a good experiment* $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}},\mathscr{A},\mu,\Phi}$*, it holds:*

*(i) No vertex* $\mathbb{V}_S$ *can be "reached" from* $\mathbb{V}_{IV}$ *via* $C+1$ *distinct prefix values. Formally, there does not exist* $C+1$ *paths* $\left(\mathbb{V}_{IV} \xrightarrow{PP_1\|\star} \mathbb{V}_S\right), ..., \left(\mathbb{V}_{IV} \xrightarrow{PP_{C+1}\|\star} \mathbb{V}_S\right)$ *that share the same endpoint* $\mathbb{V}_S$*,* $S \in \{0,1\}^c$*, but* $PP_1, ..., PP_{C+1}$ *are distinct;*

*(ii) Let* $N_k$ *be the number of vertexes that can be "reached" from* $\mathbb{V}_{IV}$ *via the k-th prefix value* $PP_k$*. Then, it holds* $\sum_{k=1}^{\alpha_P} N_k \leq CQ$*.*

*(iii) For every vertex* $\mathbb{V}_Z$*,* $Z \in \{0,1\}^h$*, the number of distinct vertexes that have limited paths to* $\mathbb{V}_Z$ *is at most* $C(\ell_{\max}-1)$*. Formally, the number of distinct* $\mathbb{V}_S$*,* $\mathbb{V}_S \neq \mathbb{V}_Z$*, such that there exists a path* $\left(\mathbb{V}_S \xrightarrow{\overline{M}} \mathbb{V}_Z\right)$ *from* $\mathbb{V}_S$ *to* $\mathbb{V}_Z$*, with* $|\overline{M}| \leq (\ell_{\max} - 1)r$*, is at most* $C(\ell_{\max} - 1)$*. Moreover, at most* $C$ *of them have indegree 0.*

*Proof.* Consider Proposition (i) first. Towards a contradiction, assume that there are $C+1$ such paths

$$\left(\mathbb{V}_{IV} \xrightarrow{PP_1\|\star} \mathbb{V}_S\right), ..., \left(\mathbb{V}_{IV} \xrightarrow{PP_{C+1}\|\star} \mathbb{V}_S\right).$$

These paths constitute a spindle-like connected component $\mathcal{C}$ in $\mathcal{G}$. Let $VV$ be the total number of vertexes in $\mathcal{C}$.

Recall that $PP_i = II\|\star$ for some user identifier value $II$, with $|II| = 2r$. Assume that depending on the involved $II$ value, $PP_1, ..., PP_{C+1}$ can be partitioned into $V_2$ disjoint sets (namely, $PP_i$ and $PP_{i'}$ are in the same set if and only if $PP_i = II\|\star$ and $PP_{i'} = II\|\star$ for the same user identifier value $II$). Further assume that: (i) these $V_2$ sets have $N_1, N_2, ..., N_{V_2}$ elements, which fulfills $\sum_{i=1}^{V_2} N_i = C+1$; (ii) the corresponding $II$ values are $II_1, ..., II_{V_2}$ respectively. Then, by Lemma 5, the number of level-2 vertexes in the connected component $\mathcal{C}$ (i.e., endpoints $S_{2,1}, ..., S_{2,V_2}$ of the paths $\left(\mathbb{V}_{IV} \xrightarrow{II_1} \mathbb{V}_{S_{2,1}}\right), ..., \left(\mathbb{V}_{IV} \xrightarrow{II_{V_2}} \mathbb{V}_{S_{2,V_2}}\right)$) must be $V_2$.

Assume that the number of level-1 vertexes in $\mathcal{C}$ (i.e., endpoints of the paths $\left(\mathbb{V}_{IV} \xrightarrow{\mathsf{left}_r(II_1)} \mathbb{V}_{S_{1,1}}\right), ..., \left(\mathbb{V}_{IV} \xrightarrow{\mathsf{left}_r(II_{V_2})} \mathbb{V}_{S_{1,V_2}}\right)$) is $V_1$. It can be proven that the total outdegree of the vertexs in $\mathcal{C}$ is at least $VV + C - 1$:

(i) The outdegree of the "source" $\mathbb{V}_{IV}$ is $V_1$;
(ii) The total outdegree of the $V_1$ level-1 vertexes must be $V_2$;
(iii) It is easy to see the total outdegree of the level-2 vertexes is at least $\sum_{i=1}^{V_2} N_i = C+1$;
(iv) The outdegree of the "destination" $\mathbb{V}_S$ is 0;
(v) For any of the other $VV - V_1 - V_2 - 2$ vertex, the outdegree is at least 1.

The total outdegree thus sums to at least $V_1 + V_2 + C + 1 + (VV - V_1 - V_2 - 2) \geq VV + C - 1$. This means the total indegree of the vertexes in $\mathcal{C}$ is at least $VV + C - 1$ as well. Conditioned on $\neg\mathsf{HitIV}$, the indegree of $\mathbb{V}_{IV}$ is 0; conditioned on $\neg\mathsf{3coll}$, every vertex has indegree at most 2. By these, it can be seen the number of vertexes of indegree 2 in $\mathcal{C}$ must be at least $C$: this indicates the occurrence of $\mathsf{MColPair}$ and contradicts the goodness of $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}^{\mathbf{\Pi},\mathsf{pd}},\mathscr{A},\mu,\Phi}$. These complete the proof of Proposition (i).

The second claim Proposition (iii) is a straightforward corollary of Proposition (i). Concretely, the total number of vertexes is at most $Q$, since each table entry $T(X) = Y$ contributes at most 1 vertex $\mathbb{V}_S$, $S = \mathsf{right}_c(X \oplus Y)$. By Proposition (i), each vertex can be "reached" from $\mathbb{V}_{IV}$ via at most $C$ distinct prefix values. By this, $\sum_{k=1}^{\alpha_P} N_k \leq CQ$.

We finally prove Proposition (iii). Consider picking $\lambda$ distinct vertexes $\mathbb{V}_{S_1}, ..., \mathbb{V}_{S_\lambda}$ in $\mathcal{G}$, such that: (i) each $\mathbb{V}_{S_i}$ has a limited path $\left(\mathbb{V}_{S_i} \xrightarrow{tail_i} \mathbb{V}_Z\right)$, and (ii) for any distinct $i_1 \neq i_2$, $\mathbb{V}_{S_{i_1}}$ and $\mathbb{V}_{S_{i_2}}$ are not in the same directed path.

We first prove that $\lambda \leq C$. For this, note that for any $i_1 \neq i_2$, since $\mathbb{V}_{S_{i_1}}$ and $\mathbb{V}_{S_{i_2}}$ both have paths to $\mathbb{V}_Z$ while $\mathbb{V}_{S_{i_1}}$ and $\mathbb{V}_{S_{i_2}}$ are not in the same path, there necessarily exists a vertex $\mathbb{V}_{S_{i_1,i_2}}$ in $\mathcal{G}$ with indegree 2, such that the paths $\left(\mathbb{V}_{S_{i_1}} \xrightarrow{tail_{i_1}} \mathbb{V}_Z\right)$ and $\left(\mathbb{V}_{S_{i_2}} \xrightarrow{tail_{i_2}} \mathbb{V}_Z\right)$ "merge" at $\mathbb{V}_{S_{i_1,i_2}}$. Conditioned on $\neg$3coll, three distinct paths cannot merge at the same vertex. By these, the $\lambda$ distinct vertexes $\mathbb{V}_{S_1}, \mathbb{V}_{S_2}, ..., \mathbb{V}_{S_\lambda}$ necessarily pinpoint at least $\lambda - 1$ distinct indegree-2 vertexes in $\mathcal{G}$. The claim $\lambda \leq C$ then follows by $\neg$MColPair.

Then, for each of these vertexes $\mathbb{V}_{S_i}$, since its path $\left(\mathbb{V}_{S_i} \xrightarrow{tail_i} \mathbb{V}_Z\right)$ is limited and is of length at most $\ell_{\max} - 1$, the number of vertexes in this path is at most $\ell_{\max} - 1$ (excluding $\mathbb{V}_Z$). Summing over the $\lambda \leq C$ vertexes $\mathbb{V}_{S_1}, ..., \mathbb{V}_{S_\lambda}$ yields the final bound $C(\ell_{\max} - 1)$.

On the other hand, for each of these $\mathbb{V}_{S_i}$, its path $\left(\mathbb{V}_{S_i} \xrightarrow{tail_i} \mathbb{V}_Z\right)$ "contributes" at most 1 vertex with indegree 0, and this yields the bound $C$ on such vertexes. □

**Number of length-7 paths.** The number of certain forms of length-7 paths will be crucial for Lemma 10, which bounds the probability of the success condition 6.a in Fig. 4.

**Lemma 7.** *In a good experiment* $\mathbf{Exp}^{\text{lms-mfpspr}}_{\text{Sponge-F}_h^{\Pi,\text{apd}},\mathscr{A},\mu,\Phi}$*, the number of length-7 paths of the form* $\left(\mathbb{V}_{IV} \xrightarrow{PP_k \| \star} \mathbb{V}_Z\right)$ *is at most* $32Q$.

*Proof.* By the constraints $\Phi$, for every $k$ it has $PP_k = II_i \| tail_k$ for some $i \in \{1, ..., \alpha_I\}$ and $|tail_k| < r$. Let $\left(\mathbb{V}_{IV} \xrightarrow{II_i} \mathbb{V}_{S_i^{II}}\right)$ be the path due to processing $II_i$. Conditioned on $\neg$ICol, the $\alpha_I$ user identifiers $II_1, ..., II_{\alpha_I}$ yield $\alpha_I$ distinct "level-2" vertexes $\mathbb{V}_{S_{2,1}} = \mathbb{V}_{S_1^{II}}, ..., \mathbb{V}_{S_{2,\alpha_I}} = \mathbb{V}_{S_{\alpha_I}^{II}}$ (they are "level-2" because the paths from $\mathbb{V}_{IV}$ to them are of length-2).

For $\ell = 3, 4, 5, 6, 7$, let $\mathbb{V}_{S_{\ell,1}}, ..., \mathbb{V}_{S_{\ell,N_\ell}}$ be the vertexes such that there exists a length-$\ell$ path of the form $\left(\mathbb{V}_{IV} \xrightarrow{PP_k \| \star} \mathbb{V}_{S_{\ell,i}}\right)$ for all $\mathbb{V}_{S_{\ell,i}}$. In some sense, this means for $\ell \in \{3, 4, 5, 6, 7\}$, there are $N_\ell$ distinct "level-$\ell$" vertexes $\mathbb{V}_{S_{\ell,1}}, ..., \mathbb{V}_{S_{\ell,N_\ell}}$. Note that vertexes at different levels may overlap: this does not affect our subsequent argument. As discussed, the number of "level-2" vertexes has $N_2 = \alpha_I$.

For $\ell \in \{2, 3, 4, 5, 6\}$ and $i \in \{1, ..., N_\ell\}$, let $L(\ell, i)$ be the number of length-$(7 - \ell)$ paths from $\mathbb{V}_{S_{\ell,i}}$ to the "level-7" vertexes $\mathbb{V}_{S_{7,1}}, ..., \mathbb{V}_{S_{7,N_7}}$. Conditioned on $\neg$ICol, each "level-2" vertex $\mathbb{V}_{S_{2,i}}$ has a unique path from $\mathbb{V}_{IV}$ to $\mathbb{V}_{S_{2,i}}$. By this, it can be seen that the number of length-7 paths of the form $\left(\mathbb{V}_{IV} \xrightarrow{PP_k \| \star} \mathbb{V}_{S_{7,i}}\right)$ exactly equals $\sum_{i=1}^{N_2} L(2, i)$, the total number of length-5 paths from "level-2" vertexes to "level-7" vertexes.

We proceed to prove:

(i) $\sum_{i=1}^{N_\ell} L(\ell, i) \leq \sum_{i=1}^{N_{\ell+1}} 2L(\ell + 1, i)$ for all $\ell \in \{2, ..., 5\}$;
(ii) $\sum_{i=1}^{N_6} L(6, i) \leq 2Q$.

Gathering the above yields that the number of length-7 paths of the form $\left(\mathbb{V}_{IV} \xrightarrow{PP_k \| \star} \mathbb{V}_{S_{7,i}}\right)$ is $\sum_{i=1}^{N_2} L(2, i) \leq 2^5 Q = 32Q$, as claimed.

We now prove Claim (i). Conditioned on $\neg$3coll, none of the "level-$(\ell+1)$" vertexes $\mathbb{V}_{S_{\ell+1,1}}, ..., \mathbb{V}_{S_{\ell+1,N_{\ell+1}}}$ has indegree $\geq 3$. Therefore, each "level-$(\ell + 1)$" vertex $\mathbb{V}_{S_{\ell+1,i}}$ falls into two cases:

- Case 1: the indegree of $\mathbb{V}_{S_{\ell+1,i}}$ is 1. Assume that $\mathbb{V}_{S_{\ell,j}}$ is the "level-$\ell$" vertex that has the edge to $\mathbb{V}_{S_{\ell+1,i}}$. Then, $\mathbb{V}_{S_{\ell+1,i}}$ "contributes" an increment of $L(\ell + 1, i)$ to the term $L(\ell, j)$, while "contributes" 0 to the other terms of the form $L(\ell, j')$. This means $\mathbb{V}_{S_{\ell+1,i}}$ "contributes" exactly $L(\ell + 1, i)$ to the sum $\sum_{i=1}^{N_\ell} L(\ell, i)$;
- Case 2: the indegree of $\mathbb{V}_{S_{\ell+1,i}}$ is 2. It further has two subcases:
  - Subcase 2.1: $\mathbb{V}_{S_{\ell,j_1}}$ and $\mathbb{V}_{S_{\ell,j_2}}$ are two distinct "level-$\ell$" vertexes, and each of them has an edge to $\mathbb{V}_{S_{\ell+1,i}}$. In this subcase, $\mathbb{V}_{S_{\ell+1,i}}$ "contributes" an increment of $L(\ell+1, i)$ to the term $L(\ell, j_1)$, $L(\ell+1, i)$ to the term $L(\ell, j_2)$, while "contributes" 0 to the other terms of the form $L(\ell, j')$;
  - Subcase 2.2: $\mathbb{V}_{S_{\ell,j}}$ is a "level-$\ell$" vertex that has 2 edges to $\mathbb{V}_{S_{\ell+1,i}}$. In this subcase, $\mathbb{V}_{S_{\ell+1,i}}$ "contributes" $2L(\ell + 1, i)$ to the term $L(\ell, j)$, while "contributes" 0 to the other terms of the form $L(\ell, j')$.

  In both subcases, $\mathbb{V}_{S_{\ell+1,i}}$ "contributes" exactly $2L(\ell + 1, i)$ to the sum $\sum_{i=1}^{N_\ell} L(\ell, i)$.

Therefore, each "level-$(\ell+1)$" vertex $\mathbb{V}_{S_{\ell+1,i}}$ "contributes" at most $2L(\ell+1, i)$ to $\sum_{i=1}^{N_\ell} L(\ell, i)$. This proves $\sum_{i=1}^{N_\ell} L(\ell, i) \leq \sum_{i=1}^{N_{\ell+1}} 2L(\ell + 1, i)$.

We finally prove Claim (ii) to conclude. The proof bears resemblance with Claim (i). Conditioned on $\neg$3coll, each "level-7" vertex $\mathbb{V}_{S_{7,i}}$ falls into two cases:

– Case 1: the indegree of $\mathbb{V}_{S_{7,i}}$ is 1. Assume that $\mathbb{V}_{S_{6,j}}$ is the "level-6" vertex that has the edge to $\mathbb{V}_{S_{7,i}}$. Then, $\mathbb{V}_{S_{7,i}}$ increases $L(6,j)$ by 1 and does not affect the other $L(6,j')$. This means $\mathbb{V}_{S_{7,i}}$ increases $\sum_{i=1}^{N_6} L(6,i)$ by 1;
– Case 2: the indegree of $\mathbb{V}_{S_{7,i}}$ is 2. It further has two subcases:
  - Subcase 2.1: $\mathbb{V}_{S_{6,j_1}}$ and $\mathbb{V}_{S_{6,j_2}}$ are two distinct "level-6" vertexes, and each of them has an edge to $\mathbb{V}_{S_{7,i}}$. In this subcase, $\mathbb{V}_{S_{7,i}}$ increases $L(6,j_1)$ and $L(6,j_2)$ by 1 and does not affect the other $L(6,j')$;
  - Subcase 2.2: $\mathbb{V}_{S_{6,j}}$ is a "level-$\ell$" vertex that has 2 edges to $\mathbb{V}_{S_{7,i}}$. In this subcase, $\mathbb{V}_{S_{7,i}}$ only increases $L(6,j)$ by 2.

  In both subcases, $\mathbb{V}_{S_{7,i}}$ increases $\sum_{i=1}^{N_6} L(6,i)$ by 2.

Therefore, each "level-7" vertex $\mathbb{V}_{S_{7,i}}$ increases $\sum_{i=1}^{N_6} L(6,i)$ by at most 2. Since "level-7" vertex must have non-zero indegree, the number of "level-7" vertex does not exceed $Q$. It thus holds $\sum_{i=1}^{N_6} L(6,i) \le 2Q$. □

**Events PostChSpr2 and PreChSpr2.** We now bound the probability of the main second preimage events PostChSpr2 and PreChSpr2.

**Lemma 8.** *In a good experiment* $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}_h^{\boldsymbol{\Pi},\mathsf{apd}},\mathscr{A},\mu,\Phi}$, *the probability that* PostChSpr2 *occurs is at most* $\frac{4\mu C^2\ell_{\max}Q+8hQ+2^{r+2}\mu CQ}{2^h}$.

The proof gains inspiration from Fluhrer's proof of [39, Lemma 4], but we take the target message length (bounded by $\ell_{\max}$) into consideration.

*Proof (Proof of Lemma 8).* We consider the $i$-th "post-challenge" $\boldsymbol{\Pi}$-query in experiment $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}_h^{\boldsymbol{\Pi},\mathsf{apd}},\mathscr{A},\mu,\Phi}$, and bound its probability of creating a "colliding path".

*Case 1: the $i$-th query is forward $\boldsymbol{\Pi}(X) \to Y$.* We further distinguish two subcases.

– Subcase 1.1: it has $\mathsf{right}_c(X\oplus Y) = Z^{i,j}$, and there exists a path $\left(\mathbb{V}_{IV} \xrightarrow{\mathsf{prefix}_{i,j}\|\star} \mathbb{V}_{\mathsf{right}_c(X)}\right)$ (note that this is a *necessarily condition*). Assume that there are $\alpha_1$ distinct prefix values $PP_1, ..., PP_{\alpha_1}$ such that there exist paths of the form $\left(\mathbb{V}_{IV} \xrightarrow{PP_k\|\star} \mathbb{V}_{\mathsf{right}_c(X)}\right)$ for $k \in \{1, ..., \alpha_1\}$. By Lemma 6 (i), it holds $\alpha_1 \le C$. Since $\max_{PP} \left|\left\{(i,j) : \mathsf{prefix}_{i,j} = PP\right\}\right| \le \mu$, the number of prefix index pairs $(i,j)$ such that there exists paths of the form $\left(\mathbb{V}_{IV} \xrightarrow{\mathsf{prefix}_{i,j}\|\star} \mathbb{V}_{\mathsf{right}_c(X)}\right)$ is at most $\mu C$. For each such prefix index pair $(i,j)$, the corresponding "target" is $Z^{i,j}$, and the probability to have $\mathsf{right}_c(X\oplus Y) = Z^{i,j}$ is at most $2^r/(2^b - Q) \le 2/2^h$. Taking a union bound over the at most $\mu C$ index pairs $(i,j)$, we see that Subcase 1.1 occurs with probability at most $\frac{2\mu C}{2^h}$.
– Subcase 1.2: it has $\mathsf{right}_c(X\oplus Y) = Z^{i,j}\oplus\theta$, and: (i) there is a path $\left(\mathbb{V}_{IV} \xrightarrow{\mathsf{prefix}_{i,j}\|\star} \mathbb{V}_{\mathsf{right}_c(X)}\right)$, and (ii) there exists a edge from $Z^{i,j}$ to $Z^{i,j}$ (again, these are *necessarily conditions*). In a similar vein to Subcase 1.1, we see that Subcase 1.2 occurs with probability at most $\frac{2\mu C}{2^h}$.
– Subcase 1.3: $\mathsf{right}_c(X\oplus Y) \notin \{Z^{i,j}, Z^{i,j}\oplus\theta\}$, $\mathsf{right}_c(X\oplus Y) = S$ or $\mathsf{right}_c(X\oplus Y) = S\oplus\theta$ "links" two paths $\left(\mathbb{V}_{IV} \xrightarrow{\mathsf{prefix}_{i,j}\|\star} \mathbb{V}_{\mathsf{right}_c(X)}\right)$ and $\left(\mathbb{V}_S \xrightarrow{tail} \mathbb{V}_{Z^{i,j}}\right)$—note that $\left(\mathbb{V}_S \xrightarrow{tail} \mathbb{V}_{Z^{i,j}}\right)$ must be limited (i.e., $|tail| \le (\ell_{\max} - 1)r$), as otherwise this "linking" does not yield valid hash paths. Similarly to Subcase 1.1, the number of prefix index pairs $(i,j)$ such that there exist paths of the form $\left(\mathbb{V}_{IV} \xrightarrow{\mathsf{prefix}_{i,j}\|\star} \mathbb{V}_{\mathsf{right}_c(X)}\right)$ is at most $\mu C$. For each such prefix index pair $(i,j)$, the corresponding "target" is $Z^{i,j}$; by Lemma 6 Proposition (iii), the number of choices of the limited path $\left(\mathbb{V}_S \xrightarrow{tail} \mathbb{V}_{Z^{i,j}}\right)$ is at most $C(\ell_{\max}-1)$. For each such path $\left(\mathbb{V}_S \xrightarrow{tail} \mathbb{V}_{Z^{i,j}}\right)$, the probability to have $\mathsf{right}_c(X\oplus Y) = S$ or $\mathsf{right}_c(X\oplus Y) = S\oplus\theta$ is at most $2^{r+1}/(2^b - Q) \le 4/2^h$. Taking a union bound over the at most $\mu C$ index pairs $(i,j)$ and the at most $C(\ell_{\max} - 1)$ choices of $\left(\mathbb{V}_S \xrightarrow{tail} \mathbb{V}_{Z^{i,j}}\right)$, we see that Subcase 1.2 occurs with probability at most $\frac{4\mu C^2(\ell_{\max}-1)}{2^h}$.

Summing over the three subcases and taking union bound over $\le Q$ forward queries, we see that Case 1 occurs with probability at most $Q \cdot \left(\frac{2\mu C}{2^h} + \frac{2\mu C}{2^h} + \frac{4\mu C^2(\ell_{\max}-1)}{2^h}\right) \le \frac{4\mu C^2\ell_{\max}Q}{2^h}$ (when $C \ge 1$).

*Case 2: the $i$-th query is backward $\boldsymbol{\Pi}^{-1}(Y) \to X$.* We further distinguish two subcases.

– Subcase 2.1: $\boldsymbol{\Pi}^{-1}(Y) \to X$ "links" two paths $\left(\mathbb{V}_{IV} \xrightarrow{\mathsf{prefix}_{i,j}\|\star} \mathbb{V}_{S_1}\right)$ and $\left(\mathbb{V}_{S_2} \xrightarrow{tail} \mathbb{V}_{Z^{i,j}}\right)$. It then holds $\mathsf{right}_c(X)\oplus\mathsf{right}_c(Y) \in \{S_2, S_2\oplus\theta\}$ and $\mathsf{right}_c(X) \in \{S_1, S_1\oplus\theta\}$, which implies $\mathsf{right}_c(Y) \in \{S_1\oplus S_2, S_1\oplus S_2\oplus\theta\}$. We consider three subcases as follows.
  - Subcase 2.1.1: $\mathbb{V}_{S_1} = \mathbb{V}_{S_2}$. Then it has $\mathsf{right}_c(Y) \in \{0^c, \theta\}$ and $\mathsf{right}_c(X) \in \{S_1, S_1\oplus\theta\}$. Since $\mathbb{V}_{S_1}$ is the endpoint of a path $\left(\mathbb{V}_{IV} \xrightarrow{\mathsf{prefix}_{i,j}\|\star} \mathbb{V}_{S_1}\right)$, before $\boldsymbol{\Pi}^{-1}(Y) \to X$ happens, there exists a table entry $T(X') = Y'$ such that $\mathsf{right}_c(X'\oplus Y') \in \{S_1, S_1\oplus\theta\}$. This means $\mathsf{right}_c(Y) \in \{0^c, \theta\}$ and $\mathsf{right}_c(X) \in \{\mathsf{right}_c(X'\oplus Y'), \mathsf{right}_c(X'\oplus Y')\oplus\theta\}$ and contradicts ¬BadLW.
    Note that this case captures the birthday attack on Merkle-Damgård with Davies-Meyer described in Sect. I.2. We thus showed that conditioned on ¬BadLW, the attack becomes ineffective.

- Subcase 2.1.2: $\mathbb{V}_{S_1} \neq \mathbb{V}_{S_2}$, and there exists another edge $T(X_2) = Y_2$ pointing to $\mathbb{V}_{S_2}$, i.e., $\mathsf{right}_c(X_2 \oplus Y_2) \in \{S_2, S_2 \oplus \theta\}$. The path $\left(\mathbb{V}_{IV} \xrightarrow{\mathsf{prefix}_{i,j} \| \star} \mathbb{V}_{S_1}\right)$ indicates the existence of an entry $T(X_1) = Y_1$ with $\mathsf{right}_c(X_1 \oplus Y_1) \in \{S_1, S_1 \oplus \theta\}$. Since $\mathbb{V}_{S_1} \neq \mathbb{V}_{S_2}$, $T(X_1) = Y_1$ and $T(X_2) = Y_2$ are distinct entries, and $\mathsf{right}_c(Y) \in \{S_1 \oplus S_2, S_1 \oplus S_2 \oplus \theta\}$ means that $\mathsf{right}_c(X_1 \oplus Y_1 \oplus X_2 \oplus Y_2)$ falls in $\{\mathsf{right}_c(Y), \mathsf{right}_c(Y) \oplus \theta\}$. Conditioned on $\neg\mathsf{hCol}$, for any choice of $Y$, the number of possible choices of such two entries $T(X_1) = Y_1$ and $T(X_2) = Y_2$ with $\mathsf{right}_c(X_1 \oplus Y_1 \oplus X_2 \oplus Y_2) \in \{\mathsf{right}_c(Y), \mathsf{right}_c(Y) \oplus \theta\}$ is at most $h$. For each pair of them, the probability to have $\mathsf{right}_c(X) \in \{S_1, S_1 \oplus \theta\}$ is at most $4/2^h$. By this, the probability that Subcase 2.1.2 occurs is at most $4hQ/2^h$.
- Subcase 2.1.3: $\mathbb{V}_{S_1} \neq \mathbb{V}_{S_2}$, and $\mathbb{V}_{S_2}$ has indegree 0 (i.e., $\mathsf{right}_c(X' \oplus Y') \notin \{S_2, S_2 \oplus \theta\}$ for all entry $T(X') = Y'$). Note that this means $\mathbb{V}_{S_2} \neq \mathbb{V}_{Z^{i,j}}$ for all pair $(i, j)$.

  The number of paths of the form $\left(\mathbb{V}_{IV} \xrightarrow{PP \| \star} \mathbb{V}_{S_1}\right)$ is at most $Q$. For each of them, the number of pairs of indices $(i, j)$ with $\mathsf{prefix}_{i,j} = PP$ is at most $\mu$, pinpointing at most $\mu$ targets $Z^{i,j}$. By Lemma 6 (iii), for each such target $Z^{i,j}$, the number of choices of $\left(\mathbb{V}_{S_2} \xrightarrow{tail} \mathbb{V}_{Z^{i,j}}\right)$ with $\mathbb{V}_{S_2}$ of indegree 0 is at most $C$. In all, the number of proper choices of $\left(\mathbb{V}_{IV} \xrightarrow{\mathsf{prefix}_{i,j} \| \star} \mathbb{V}_{S_1}\right)$ and $\left(\mathbb{V}_{S_2} \xrightarrow{tail} \mathbb{V}_{Z^{i,j}}\right)$ is at most $\mu CQ$. For each $(\mathbb{V}_{S_1}, \mathbb{V}_{S_2})$, the condition $\mathsf{right}_c(Y) \in \{S_1 \oplus S_2, S_1 \oplus S_2 \oplus \theta\}$ restricts the number of choices of adversarial query $\mathbf{\Pi}(Y) \to X$ to at most $2^{r+1}$. By these, the probability that Subcase 2.1.3 occurs is at most $2^{r+1}\mu CQ \cdot \frac{2^r}{2^b - Q} \leq \frac{2^{r+2}\mu CQ}{2^b}$.

Summing over the subcases yields that Subcase 2.1 occurs with probability at most $\frac{4hQ + 2^{r+2}\mu CQ}{2^b}$.

– Subcase 2.2: $\mathbf{\Pi}^{-1}(Y) \to X$ "links" a path $\left(\mathbb{V}_{IV} \xrightarrow{\mathsf{prefix}_{i,j} \| \star} \mathbb{V}_{S_1}\right)$ to the target $Z^{i,j}$. It then holds $\mathsf{right}_c(X) \oplus \mathsf{right}_c(Y) = Z^{i,j}$ and $\mathsf{right}_c(X) \in \{S_1, S_1 \oplus \theta\}$, which implies $\mathsf{right}_c(Y) \in \{S_1 \oplus Z^{i,j}, S_1 \oplus Z^{i,j} \oplus \theta\}$. We consider two subcases as follows.
  - Subcase 2.3.1: $\mathbb{V}_{S_1} = \mathbb{V}_{Z^{i,j}}$. Then it has $\mathsf{right}_c(Y) \in \{0^c, \theta\}$ and $\mathsf{right}_c(X) \in \{S_1, S_1 \oplus \theta\}$. Since $\mathbb{V}_{S_1}$ is endpoint of $\left(\mathbb{V}_{IV} \xrightarrow{\mathsf{prefix}_{i,j} \| \star} \mathbb{V}_{S_1}\right)$, before $\mathbf{\Pi}^{-1}(Y) \to X$ happens, there exists a table entry $T(X') = Y'$ such that $\mathsf{right}_c(X' \oplus Y') \in \{S_1, S_1 \oplus \theta\}$. This means $\mathsf{right}_c(Y) \in \{0^c, \theta\}$ and $\mathsf{right}_c(X)$ falls in the set $\{\mathsf{right}_c(X' \oplus Y'), \mathsf{right}_c(X' \oplus Y') \oplus \theta\}$ and contradicts $\neg\mathsf{BadLW}$.
  - Subcase 2.3.2: $\mathbb{V}_{S_1} \neq \mathbb{V}_{Z^{i,j}}$. The vertex $\mathbb{V}_{Z^{i,j}}$ certainly has indegree $\geq 1$, i.e., there exists another edge $T(X_2) = Y_2$ with $\mathsf{right}_c(X_2 \oplus Y_2) \in \{Z^{i,j}, Z^{i,j} \oplus \theta\}$. The path $\left(\mathbb{V}_{IV} \xrightarrow{\mathsf{prefix}_{i,j} \| \star} \mathbb{V}_{S_1}\right)$ indicates the existence of an entry $T(X_1) = Y_1$ with $\mathsf{right}_c(X_1 \oplus Y_1) \in \{S_1, S_1 \oplus \theta\}$. Since $\mathbb{V}_{S_1} \neq \mathbb{V}_{Z^{i,j}}$, $T(X_1) = Y_1$ and $T(X_2) = Y_2$ are distinct entries, and $\mathsf{right}_c(Y) \in \{S_1 \oplus Z^{i,j}, S_1 \oplus Z^{i,j} \oplus \theta\}$ means $\mathsf{right}_c(X_1 \oplus Y_1 \oplus X_2 \oplus Y_2) \in \{\mathsf{right}_c(Y), \mathsf{right}_c(Y) \oplus \theta\}$. Conditioned on $\neg\mathsf{hCol}$, for any choice of $Y$, the number of possible choices of such two entries $T(X_1) = Y_1$ and $T(X_2) = Y_2$ with $\mathsf{right}_c(X_1 \oplus Y_1 \oplus X_2 \oplus Y_2) \in \{\mathsf{right}_c(Y), \mathsf{right}_c(Y) \oplus \theta\}$ is at most $h$. For each pair of them, the probability to have $\mathsf{right}_c(X) \in \{S_1, S_1 \oplus \theta\}$ is at most $4/2^h$. By this, the probability that Subcase 2.3.2 occurs is at most $4hQ/2^h$.

Subcase 2.3 thus occurs with probability at most $\frac{4hQ}{2^h}$.

Summing over the two subcases, we see that Case 2 occurs with probability at most $\frac{4hQ + 2^{r+2}\mu CQ}{2^b} + \frac{4hQ}{2^h} \leq \frac{8hQ + 2^{r+2}\mu CQ}{2^b}$. The claim thus follows. □

**Lemma 9.** *In a good experiment* $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}},\mathscr{A},\mu,\Phi}$, *the probability that* PreChSpr2 *occurs is at most* $\frac{4\mu C^2 \ell_{\max}^2 Q}{2^b} + \frac{4\mu CQ}{2^h}$.

*Proof.* Consider generating the challenge $Z^{i,j} \leftarrow \mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}}(\mathsf{prefix}_{i,j} \| M^{i,j})$, which creates a path $\left(\mathbb{V}_{IV} \xrightarrow{\mathsf{prefix}_{i,j} \| M^{i,j} \| pad} \mathbb{V}_{Z^{i,j}}\right)$ in $\mathcal{G}$, $pad = sf(\mathsf{prefix}_{i,j} \| M^{i,j})$. Let $\mathsf{prefix}_{i,j} = PP_k$, and let $\mathsf{prefix}_{i,j} \| M^{i,j} \| pad = \overline{M}[1] \| ... \| \overline{M}[\ell_{i,j}]$. Recall from Lemma 6 (ii) that $N_k$ denotes the number of vertexes that can be "reached" from $IV$ via $PP_k$.

Since $\left(\mathbb{V}_{IV} \xrightarrow{\mathsf{prefix}_{i,j} \| M^{i,j} \| pad} \mathbb{V}_{Z^{i,j}}\right)$ is newly created in **Phase 2** of the experiment $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}},\mathscr{A},\mu,\Phi}$, there occurs at least 1 new forward $\mathbf{\Pi}$-queries. Let $\mathbf{\Pi}(X) \to Y$ be a new $\mathbf{\Pi}$-query occurred during creating $\left(\mathbb{V}_{IV} \xrightarrow{\mathsf{prefix}_{i,j} \| M^{i,j} \| pad} \mathbb{V}_{Z^{i,j}}\right)$, and assume that $\mathbf{\Pi}(X) \to Y$ was due to absorbing the $\ell'$-th block $\overline{M}[\ell']$, $\ell' < \ell_{i,j}$: if it "hits" some appropriate paths then PreChSpr2 may occur. Concretely, if: (i) there exist a path $\left(\mathbb{V}_{IV} \xrightarrow{PP_k \| M'} \mathbb{V}_{Z^{i,j}}\right)$, and (ii) there exist a path $\left(\mathbb{V}_S \xrightarrow{W \| \overline{M}[\ell'+2] \| ... \| \overline{M}[\ell_{i,j}]} \mathbb{V}_{Z^{i,j}}\right)$, and (iii) let $T(X_1) = Y_1, ..., T(X_{\ell_{i,j} - \ell'}) = Y_{\ell_{i,j} - \ell'}$ be the table entries underlying $\left(\mathbb{V}_S \xrightarrow{W \| \overline{M}[\ell'+2] \| ... \| \overline{M}[\ell_{i,j}]} \mathbb{V}_{Z^{i,j}}\right)$, then it holds $\mathsf{left}_r(Y) = \mathsf{left}_r(X_1) \oplus \overline{M}[\ell' + 1]$, and $\mathsf{right}_c(X \oplus Y) = \mathsf{right}_c(X_1)$ (when $\ell_{i,j} - \ell' \geq 2$) or $\mathsf{right}_c(X \oplus Y) = \mathsf{right}_c(X_1) \oplus \theta$ (when $\ell_{i,j} - \ell' = 1$).

Since the vertex $\mathbb{V}_{Z^{i,j}}$ is "reachable" from $\mathbb{V}_{IV}$ via the prefix value $PP_k$, the number of choices for $\mathbb{V}_{Z^{i,j}}$ is at most $N_k$. For each choice of $\mathbb{V}_{Z^{i,j}}$, the number of choices for $\mathbb{V}_S$ such that there exists the path $\left(\mathbb{V}_S \xrightarrow{W \| \overline{M}[\ell'+2] \| ... \| \overline{M}[\ell_{i,j}]} \mathbb{V}_{Z^{i,j}}\right)$ is at most $C(\ell_{\max} - 1)$ Lemma 6 (iii). Conditioned on $\neg\mathsf{3coll}$, each such path $\left(\mathbb{V}_S \xrightarrow{W \| \overline{M}[\ell'+2] \| ... \| \overline{M}[\ell_{i,j}]} \mathbb{V}_{Z^{i,j}}\right)$ raise at most

2 distinct choices for the first underlying entry $T(X_1) = Y_1$: for each of them the probability that condition (iii) holds is at most $1/(2^b - Q)$. By these, the probability that the new query $\mathbf{\Pi}(X) \to Y$ incurs the event $\mathsf{PreChSpr2}$ is at most $2C(\ell_{\max}-1)N_k/(2^b-Q) \le 4C\ell_{\max}N_k/2^b$. Since the number of such new $\mathbf{\Pi}$-queries due to creating $\big(\mathbb{V}_{IV} \xrightarrow{\mathsf{prefix}_{i,j}\|M^{i,j}\|pad} \mathbb{V}_{Z^{i,j}}\big)$ is at most $\ell_{\max}$, the probability is at most $4C\ell_{\max}^2 N_k/2^b$ in total.

When $\mathbf{\Pi}(X) \to Y$ was due to absorbing last block $\overline{M}[\ell_{i,j}]$, the condition is slightly simplified—if: (i) there exist a path $\big(\mathbb{V}_{IV} \xrightarrow{PP_k\|M'} \mathbb{V}_{Z^{i,j}}\big)$, and (ii) $\mathsf{right}_c(X \oplus Y) \in \{Z^{i,j}, Z^{i,j} \oplus \theta\}$, then $\mathbf{\Pi}(X) \to Y$ may incur $\mathsf{PreChSpr2}$. In a similar vein to the above, it can be seen the probability is at most $2^{r+1}N_k/(2^b - Q) \le 4N_k/2^h$. Gathering the two cases, the probability that $\mathsf{PreChSpr2}$ occurs due to $Z^{i,j} \leftarrow \mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}}(\mathsf{prefix}_{i,j}\|M^{i,j})$ is at most $4C\ell_{\max}^2 N_k/2^b + 4N_k/2^h$.

Let $\varphi(i,j)$ be the function that maps $(i,j)$ to $k \in \{1,...,\alpha_P\}$, such that $\mathsf{prefix}_{i,j} = PP_k$. Then, every $k \in \{1,...,\alpha_P\}$ has at most $\mu$ distinct preimages under $\varphi$. Summing over all index pairs $(i,j)$ then yields

$$\begin{aligned}\Pr\big[\mathsf{PreChSpr2}\big] &\le \sum_{i=1,2,3}\ \sum_{j\in\{1,\ldots,\gamma_i\}} \left(\frac{4C\ell_{\max}^2 N_{\varphi(i,j)}}{2^b} + \frac{4N_{\varphi(i,j)}}{2^h}\right) \\ &\le \mu \cdot \sum_{k=1}^{\alpha_P} \left(\frac{4C\ell_{\max}^2 N_k}{2^b} + \frac{4N_k}{2^h}\right) \\ &\le \frac{4\mu C^2 \ell_{\max}^2 Q}{2^b} + \frac{4\mu C Q}{2^h}, \end{aligned} \tag{64}$$

where the last inequality follows from $\sum_{k=1}^{\alpha_P} N_k \le CQ$ by Lemma 6 (ii). □

**Bounding attack success probability.** Conditioned on that the experiment is good and that the second preimage events $\mathsf{PostChSpr2}$ and $\mathsf{PreChSpr2}$ did not occur, the following three lemmas give bounds on adversarial success w.r.t. the three groups of challenges in turn.

**Lemma 10.** *At the end of a good experiment* $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}},\mathscr{A},\mu,\Phi}$*, if* $\mathsf{PostChSpr2}$ *never occurred, then the success condition 6.a in Fig. 4 is fulfilled with probability at most* $\frac{32\mu Q}{2^h}$*.*

Intuitively, since $\mathsf{PostChSpr2}$ did not occur, the only possibility is that $\mathscr{A}$ succeeds in "guessing" the challenge $M^{1,i} \xleftarrow{\$} \{0,1\}^h$ for some $\mathsf{prefix}_{1,i}$. The format of $\mathsf{prefix}_{1,i}$ and $h = 4r$ means that a path of length-7 formed by seven $\mathbf{\Pi}$ query records of $\mathscr{A}$ contains the full $M^{1,i}$, the probability is $1/2^h$ due to uniformness of $M^{1,i}$. The number of such length-7 paths has been bounded to $32Q$ by Lemma 7, and this yields the claimed bound.

*Proof.* Consider the event that there exists an index $i \in \{1,...,\gamma_1\}$ such that $\mathscr{A}$ finds $M'$ such that $\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}}(\mathsf{prefix}_{1,i}\|M') = \mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}}(\mathsf{prefix}_{1,i}\|M^{1,i})$. By the definition of the lms-mfpspr experiment $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}},\mathscr{A},\mu,\Phi}$ (Fig. 4), the Group-1 challenge evaluation $Z^{1,i} \leftarrow \mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}}(\mathsf{prefix}_{1,i}\|M^{1,i})$ occurs before $\mathscr{A}$ is allowed to query $\mathbf{\Pi}$. By this, conditioned on $\neg\mathsf{PostChSpr2}$, it necessarily has $M' = M^{1,i}$, i.e., $\mathscr{A}$ succeeds in "guessing" the value of $M^{1,i}$. This means $M^{1,i}$ "appears" in the $\mathbf{\Pi}$-queries of $\mathscr{A}$.

Consider organizing $\mathscr{A}$'s $\mathbf{\Pi}$-queries and responses as *paths*, which resemble the aforementioned form $\big(\mathbb{V}_{IV} \xrightarrow{\overline{M}} \mathbb{V}_S\big)$. Then, $M^{1,i}$ "appears" in the $\mathbf{\Pi}$-queries of $\mathscr{A}$ means that there appears a path of the form $\big(\mathbb{V}_{IV} \xrightarrow{PP_k\|M^{1,i}\|pad} \mathbb{V}_S\big)$ for some $i \in \{1,...,\gamma_1\}$, where $PP_k = \mathsf{prefix}_{1,i}$ and $pad = sf(PP_k\|M^{1,i})$. For convenience, denote this event by $\mathsf{BadGuess}$.

Note that $|PP_k\|M^{1,i}\|pad| = 7r$, where $M^{1,i} \xleftarrow{\$} \{0,1\}^h$ is uniformly picked. By this, for any $M$ with $|M| = h = 4r$, the path $\big(\mathbb{V}_{IV} \xrightarrow{PP_k\|M\|pad} \mathbb{V}_S\big)$, $pad = sf(PP_k\|M)$, is of length 7. For each such length-7 path $\big(\mathbb{V}_{IV} \xrightarrow{PP_k\|M\|pad} \mathbb{V}_S\big)$, the probability to have $M^{1,i} = M$ (i.e., $\mathscr{A}$ "guesses" $M^{1,i}$) is $1/2^h$. By this, let $NN_k$ be the number of length-7 paths of the form $\big(\mathbb{V}_{IV} \xrightarrow{PP_k\|M\|pad} \mathbb{V}_S\big)$. Then, the probability that $M^{1,i}$ "appears" in the $\mathbf{\Pi}$-queries of $\mathscr{A}$ is at most $NN_k/2^h$.

Let $\varphi(i,j)$ be the function that maps $(i,j)$ to $k \in \{1,...,\alpha_P\}$, such that $\mathsf{prefix}_{i,j} = PP_k$. Then, summing over all index pairs $(i,1)$ yields

$$\Pr\big[\mathsf{BadGuess}\big] \le \sum_{i\in\{1,\ldots,\gamma_1\}} \frac{NN_{\varphi(1,i)}}{2^h} \le \mu \cdot \sum_{k=1}^{\alpha_P} \frac{NN_k}{2^h} \le \frac{32\mu Q}{2^h}, \tag{65}$$

where the last inequality follows from $\sum_{k=1}^{\alpha_P} NN_k \le 32Q$ by Lemma 7 (the number of length-7 paths of the form $\big(\mathbb{V}_{IV} \xrightarrow{PP_k\|M\|pad} \mathbb{V}_S\big)$ is at most $32Q$). □

**Lemma 11.** *At the end of a good experiment* $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}},\mathscr{A},\mu,\Phi}$*, if* $\mathsf{PostChSpr2}$ *never occurred, then the success condition 6.b in Fig. 4 cannot be fulfilled.*

*Proof.* By the definition of the lms-mfpspr experiment $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}},\mathscr{A},\mu,\Phi}$ (Fig. 4), the Group-2 challenge evaluation $Z^{2,i} \leftarrow \mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}}(\mathsf{prefix}_{2,i}\|M^{2,i})$ occurs before $\mathscr{A}$ is allowed to query $\mathbf{\Pi}$. The claim is then obvious by the definition of PostChSpr2. □

**Lemma 12.** *At the end of a good* $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}},\mathscr{A},\mu,\Phi}$*, if neither* PostChSpr2 *nor* PreChSpr2 *occurred, then the success condition 6.c in Fig. 4 is fulfilled with probability at most* $\frac{32\mu Q}{2^h}$.

Intuitively, since neither PostChSpr2 nor PreChSpr2 occurred, the only possibility is that some challenge randomness $R^i$ is "bad" such that $\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}}(\mathsf{prefix}_{3,i}\|R^i\|M^{3,i})$ does not yield a new path, the probability of which is the same as Lemma 10.

*Proof.* Consider the event that there exists an index $i \in \{1,...,\gamma_3\}$ such that $\mathscr{A}$ finds $M'$ such that $\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}}(\mathsf{prefix}_{3,i}\|M') = Z^{3,i} = \mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}}(\mathsf{prefix}_{3,i}\|R^i\|M^{3,i})$. Conditioned on ¬PostChSpr2, $M'$ cannot be found by $\mathscr{A}$ *after* the challenges $R^i$ and $Z^{3,i}$ are generated. Conditioned on ¬PreChSpr2, if evaluating the challenge $Z^{3,i} \leftarrow \mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}}(\mathsf{prefix}_{3,i}\|R^i\|M^{3,i})$ results in a new path $\big(\mathbb{V}_{IV} \xrightarrow{\mathsf{prefix}_{3,i}\|R^i\|M^{3,i}\|pad} \mathbb{V}_{Z^{3,i}}\big)$, $pad = sf(\mathsf{prefix}_{3,i}\|R^i\|M^{3,i})$, then $\mathscr{A}$ cannot find $M'$ from early-created paths either. This means the only possibility is that the randomness $R^i$ is not "good", such that $\mathsf{prefix}_{3,i}\|R^i\|M^{3,i}\|pad$ "hits" an existing path. More formally, let G3BadR be the event that there exists a Group-3 index $i \in \{1,...,\gamma_3\}$ such that after the $h$-bit challenge random string $R^i$ is sampled from $\{0,1\}^h$, there already exists a length-7 path $\big(\mathbb{V}_{IV} \xrightarrow{\mathsf{prefix}_{3,i}\|R^i\|\mathsf{left}_a(M^{3,i})} \mathbb{V}_S\big)$ (for some appropriate integer $a$). Since $R^i$ is uniformly picked from $\{0,1\}^h$, the event G3BadR resembles the success condition 6.a in Fig. 4: following the same line as the proof of Lemma 10, it can be shown

$$\Pr\big[\mathsf{G3BadR}\big] \le \frac{32\mu Q}{2^h}. \tag{66}$$

Thus the claim. □

Finally, gathering the bounds in Eq. (63) and Lemmas 8, 9, 10, 11 and 12 yields the main claim:

$$\begin{aligned}
\mathbf{Adv}^{\text{lms-mfpspr}[\Phi]}_{\mathsf{Sponge\text{-}F}_h^{\mathbf{\Pi},\mathsf{apd}}}(\mathscr{A}) &\le \frac{8Q + 2^{r+2}Q + 4\mu C^2\ell_{\max}Q + 8u^2 + 8hQ + 2^{r+2}\mu CQ + 64\mu Q}{2^h} \\
&\quad + \frac{4Q^3}{2^{2h}} + \Big(\frac{2Q^2}{2^h}\Big)^C \cdot \frac{1}{C!} + \frac{8Q^2}{h!2^h} + \frac{4\mu C^2\ell_{\max}^2 Q}{2^b} + \frac{4\mu CQ}{2^h} \\
&\le \frac{8u^2 + 4\mu C^2\ell_{\max}Q + 2^{r+2}(\mu C+1)Q + 84\mu ChQ}{2^h} \\
&\quad + \frac{4Q^3}{2^{2h}} + \Big(\frac{2Q^2}{2^h}\Big)^C \cdot \frac{1}{C!} + \frac{8Q^2}{h!2^h} + \frac{4\mu C^2\ell_{\max}^2 Q}{2^b}.
\end{aligned} \tag{67}$$

# G Proof of Theorem 6 (Indifferentiability of $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$)

This section is organized as follows. Sect. G.1 introduces the simulator used for the proof, Sect. G.2 shows proof overview, Sect. G.3 bounds simulator complexities, Sect. G.4 bounds the probability of simulator abortions, Sect. G.5 proves consistency of the simulation, and Sect. G.6 finally proves indistinguishability of the real and ideal worlds to complete the proof.

## G.1 Simulator Definition

**Internal variables.** The simulator $\mathcal{S}^{\mathbf{VRO}}$ maintains a pair of tables $T$ and $T^{-1}$ that map entries $X \in \{0,1\}^b$ to $Y \in \{0,1\}^b$, and that are initially empty. They denote inputs/outputs of $\mathcal{S}^{\mathbf{VRO}}.\mathrm{P}$ and $\mathcal{S}^{\mathbf{VRO}}.\mathrm{P}^{-1}$. $\mathcal{S}^{\mathbf{VRO}}$ maintains consistency in $T$ and $T^{-1}$, so that they always define a partial permutation: for any $(X,Y) \in (\{0,1\}^b)^2$, $T(X) = Y$ if and only if $T^{-1}(Y) = X$. For this, once $\mathcal{S}^{\mathbf{VRO}}$ is forced to write inconsistent entries to $T$ and $T^{-1}$, it *aborts*.

A sequence of table entries that fit into the computation of $\mathsf{Sponge\text{-}F}^{\mathbf{\Pi},\mathsf{pd}}(M)$ for some $M$ form a *(computation) path*. This includes two cases:

– Case 1: a sequence of table entries

$$T(X_1) = Y_1,\ T(X_2) = Y_2, ...,\ T(X_\ell) = Y_\ell$$

satisfying $X_1 = \overline{M}[1]\|IV$ for some $\overline{M}[1] \in \{0,1\}^r$ and $\mathsf{right}_c(X_{i+1}) = \mathsf{right}_c(X_i \oplus Y_i)$ for all $i = 1, ..., \ell-1$ may be the first $\ell$ permutation calls (in particular, *before* the XOR of $\theta$) of the computation of $\mathsf{Sponge\text{-}F}^{T,\mathsf{pd}}(M,\nu)$ for some $(M,\nu)$, where $\mathsf{pd}(M) = \overline{M}[1]\|...\|\overline{M}[\ell]\|..., \overline{M}[i+1] = \mathsf{left}_r(Y_i) \oplus \mathsf{left}_r(X_{i+1})$ for $i = 1, ..., \ell-1$. We call this an *inner (computation) path and denote* $(\overline{M}[1]\|...\|\overline{M}[\ell], SC, 0)$, where $SC = \big(\mathsf{left}_r(Y_\ell)\|\mathsf{right}_c(X_\ell \oplus Y_\ell)\big)$.

– Case 2: a sequence of table entries

$$T(X_1) = Y_1,\ T(X_2) = Y_2, ...,\ T(X_\ell) = Y_\ell$$

such that there exists $j \in \{1, ..., \ell\}$ such that:
(i) $X_1 = \overline{M}[1] \| IV$ (when $j < \ell$) or $X_1 = \overline{M}[1] \| (IV \oplus \theta)$ (when $j = \ell$) for some $\overline{M}[1] \in \{0,1\}^r$;
(ii) $\mathsf{right}_c(X_{i+1}) = \mathsf{right}_c(X_i \oplus Y_i)$ for all $i = 1, ..., \ell - j - 1, \ell - j + 1, ..., \ell - 1$;
(iii) $\mathsf{right}_c(X_{\ell-j+1}) = \mathsf{right}_c(X_{\ell-j} \oplus Y_{\ell-j}) \oplus \theta$;
may be the permutation calls of the computation of $\mathsf{Sponge}\text{-}\mathsf{F}^{T,\mathsf{pd}}(M, jr')$ for some $M$, where $\mathsf{pd}(M) = \overline{M}[1] \| ... \| \overline{M}[\ell - j + 1]$, $\overline{M}[i] = \mathsf{left}_r(Y_{i-1}) \oplus \mathsf{left}_r(X_i)$ for $i = 2, ..., \ell - j + 1$. We call this an *outer (computation) path and denote* $(\overline{M}[1] \| ... \| \overline{M}[\ell - j + 1], SC, j)$, where $SC = \big(\mathsf{left}_r(Y_\ell) \| \mathsf{right}_c(X_\ell \oplus Y_\ell)\big)$. Note that in this second case, the padded message must have $\mathsf{unpd}(\overline{M}[1] \| ... \| \overline{M}[\ell - j + 1]) \neq \perp$.
In such an outer path, the first $\ell - j$ blocks $T(X_1) = Y_1, ...,\ T(X_{\ell-j}) = Y_{\ell-j}$ also form an inner path $(\overline{M}[1] \| ... \| \overline{M}[\ell - j], SC', 0)$, where the final state has $SC' = \big(\mathsf{left}_r(Y_{\ell-j}) \| \mathsf{right}_c(X_{\ell-j} \oplus Y_{\ell-j})\big)$. We call it *the inner part* of the outer path $(\overline{M}[1] \| ... \| \overline{M}[\ell - j + 1], SC, j)$. As will be seen, the inner path $(\overline{M}[1] \| ... \| \overline{M}[\ell - j], SC', 0)$ will also be recorded by our simulator.

$\mathcal{S}^{\mathbf{VRO}}$ maintains a set *Paths* to keep track of such paths. Entries in *Paths* are of the form $(\overline{M}, SC, j)$, where $\overline{M}$ is the padded message reflected by this path, $SC \in \{0,1\}^b$ is the $b$-bit final state of this path, and $j \in \mathbb{N}_0$ is a natural number: when $j = 0$, it denotes an inner path; when $j \geq 1$, it denotes an outer path that has encountered $j$ squeezing actions. As will be shown in the next subsubsection, $\mathcal{S}^{\mathbf{VRO}}$ will *detect* and maintain inner paths and *complete* them into outer paths to ensure consistency with **VRO**.

**Simulator strategy.** The simulator $\mathcal{S}^{\mathbf{VRO}}$ is described using pseudocode in Figs. 8 and 9: the former presents variables and simulated interface P, while the latter presents the simulated interface $\mathrm{P}^{-1}$.

Briefly speaking, upon an adversarial forward query $\mathrm{P}(X)$, if the table entry $T(X)$ has been defined then $\mathcal{S}^{\mathbf{VRO}}$ simply returns $T(X)$. Otherwise, $\mathcal{S}^{\mathbf{VRO}}$ reacts depending on a case-study.

– If there is an inner path $(\overline{M}^*, SC^*, 0) \in \mathit{Paths}$ such that $\mathsf{right}_c(X) = \mathsf{right}_c(SC^*) \oplus \theta$ and an unpadded message $M$ can be correctly computed from $(\overline{M}^*, SC^*, 0)$ and $X$, then $\mathcal{S}^{\mathbf{VRO}}$ defines $T(X)$ to be consistent with **VRO**. Concretely, $\mathcal{S}^{\mathbf{VRO}}$ computes $m \leftarrow \mathsf{left}_r(SC^* \oplus X)$ and $M \leftarrow \mathsf{unpd}(\overline{M}^* \| m)$, queries $Z[1] \leftarrow \mathbf{VRO}(M, r')$, samples $y_1 \xleftarrow{\$} \{0,1\}^{b-r'}$ and defines $Y \leftarrow (y_1 \| Z[1]) \oplus (0^r \| \mathsf{right}_c(X))$. $\mathcal{S}^{\mathbf{VRO}}$ then adds an outer path $(\overline{M}, y_1 \| Z[1], 1)$ into *Paths*. We refer to Fig. 8, lines 8 to 18 for details.
– If there exists an outer path $(\overline{M}, SC^*, j) \in \mathit{Paths}$, $j \geq 1$, such that $\mathsf{right}_c(SC^*) = \mathsf{right}_c(X)$, then $\mathcal{S}^{\mathbf{VRO}}$ defines $T(X)$ to be consistent with **VRO**. Concretely, $\mathcal{S}^{\mathbf{VRO}}$ queries $Z[j+1] \leftarrow \mathsf{right}_{r'}\big(\mathbf{VRO}(M, (j+1)r')\big)$, samples $y_1 \xleftarrow{\$} \{0,1\}^{b-r'}$ and defines $Y \leftarrow (y_1 \| Z[j+1]) \oplus (0^r \| \mathsf{right}_c(X))$. $\mathcal{S}^{\mathbf{VRO}}$ then adds a new outer path $(\overline{M}, y_1 \| Z[j+1], j+1)$ into *Paths*. We refer to Fig. 8, lines 19 to 27 for details.
– Otherwise, i.e., $\mathrm{P}(X)$ neither forms new outer path nor "extends" existing outer paths, then $\mathcal{S}^{\mathbf{VRO}}$ simply samples $Y \xleftarrow{\$} \{0,1\}^b$ as the answer.

In all the above cases, $\mathcal{S}^{\mathbf{VRO}}$ checks if the to-be-defined value $Y$ breaks consistency in $(T, T^{-1})$, i.e., $Y \in T^{-1}$, and aborts if so, while defines $T(X) \leftarrow Y$ and $T^{-1}(Y) \leftarrow X$ otherwise. We refer to Fig. 8, line 31 for details. After these, $\mathcal{S}^{\mathbf{VRO}}$ returns $T(X)$ as the answer to $\mathrm{P}(X)$.

The reactions to reply an adversarial backward query $\mathrm{P}^{-1}(Y)$ are much simpler. Concretely, $\mathcal{S}^{\mathbf{VRO}}$ returns $T^{-1}(Y)$ if $T^{-1}(Y)$ has been defined. Otherwise, $\mathcal{S}^{\mathbf{VRO}}$ samples an input $X \xleftarrow{\$} \{0,1\}^b$, aborts if $X \in T$, and defines $T(X) \leftarrow Y$ and $T^{-1}(Y) \leftarrow X$ otherwise. We refer to Fig. 9 for details. $\mathcal{S}^{\mathbf{VRO}}$ finally returns $T^{-1}(Y)$ as the answer to $\mathrm{P}^{-1}(Y)$.

## G.2 Outline of the Proof

Let $\Sigma_{\mathrm{id}}$ and $\Sigma_{\mathrm{re}}$ be the ideal world and real world oracles, i.e.,

– Ideal world:

$$\Sigma_{\mathrm{id}} = (\mathrm{P}, \mathrm{P}^{-1}, \mathrm{VOLH}) = (\mathcal{S}^{\mathbf{VRO}}.\mathrm{P}, \mathcal{S}^{\mathbf{VRO}}.\mathrm{P}^{-1}, \mathbf{VRO});$$

– Real world:

$$\Sigma_{\mathrm{re}} = (\mathrm{P}, \mathrm{P}^{-1}, \mathrm{VOLH}) = (\mathbf{\Pi}, \mathbf{\Pi}^{-1}, \mathsf{Sponge}\text{-}\mathsf{F}^{\mathbf{\Pi},\mathsf{pd}}).$$

**Variables**: tables $T$ and $T^{-1}$, initially empty
Sets $Paths$, initialized to $\{(\text{empty\_string}, 0^r \| IV, 0)\}$
**public procedure** $\mathrm{P}(X)$
**if** $X \in T$ **then**
  **return** $T(X)$
**if** $\exists$ distinct $(\overline{M}, SC, j), (\overline{M}', SC', j') \in Paths$ such that $\mathsf{right}_c(SC) = \mathsf{right}_c(SC')$ or $\mathsf{right}_c(SC) = \mathsf{right}_c(SC') \oplus \theta$ **then**
  **abort**
**if** $\exists\ (\overline{M}^*, SC^*, 0) \in Paths$ such that $\mathsf{right}_c(SC^*) \oplus \theta = \mathsf{right}_c(X)$ **then**
  // Note that this includes the case $X = \star \| (IV \oplus \theta)$
  $m \leftarrow \mathsf{left}_r(SC^*) \oplus \mathsf{left}_r(X)$
  $\overline{M} \leftarrow \overline{M}^* \| m$, $M \leftarrow \mathsf{unpd}(\overline{M})$
  **if** $M \neq \perp$ **then**
    $Z[1] \leftarrow \mathbf{VRO}(M, r')$, $y_1 \xleftarrow{\$} \{0,1\}^{b-r'}$, $SC \leftarrow y_1 \| Z[1]$
    $Y \leftarrow \big((0^r \| \mathsf{right}_c(SC^*)) \oplus SC\big)$
    **if** $Y \in T^{-1}$ **then**
      **abort**
    $T(X) \leftarrow Y$, $T^{-1}(Y) \leftarrow X$
    Adds $(\overline{M}, SC, 1)$ to $Paths$
**else if** $\exists\ (\overline{M}, SC^*, j) \in Paths$ such that $j \geq 1$ and $SC^* = X$ **then**// Extend
  $M \leftarrow \mathsf{unpd}(\overline{M})$
  $Z[j+1] \leftarrow \mathsf{right}_{r'}\big(\mathbf{VRO}(M, (j+1)r')\big)$
  $y_1 \xleftarrow{\$} \{0,1\}^{b-r'}$, $SC \leftarrow y_1 \| Z[j+1]$
  $Y \leftarrow \big((0^r \| \mathsf{right}_c(SC^*)) \oplus SC\big)$
  **if** $Y \in T^{-1}$ **then**
    **abort**
  $T(X) \leftarrow Y$, $T^{-1}(Y) \leftarrow X$
  Adds $(M, SC, j+1)$ to $Paths$
// If $T(X)$ is not defined by the above then call RandAssign($X$).
**if** $X \notin T$ **then**
  $Y \xleftarrow{\$} \{0,1\}^b$, $SC \leftarrow \mathsf{left}_r(Y) \| \mathsf{right}_c(X \oplus Y)$ // $SC \in \{0,1\}^b$
  **if** $Y' \in T^{-1}$ **then**
    **abort**
  $T(X) \leftarrow Y$, $T^{-1}(Y) \leftarrow X$
**if** $\exists\ (\overline{M}^*, SC^*, 0) \in Paths$ such that $\mathsf{right}_c(SC^*) = \mathsf{right}_c(X)$ **then**
  // Extend the path: note that this includes the case $X = \star \| IV$
  $m \leftarrow \mathsf{left}_r(SC^*) \oplus \mathsf{left}_r(X)$
  Add $(\overline{M}^* \| m, Y \oplus (0^r \| \mathsf{right}_c(X)))$ to $Paths$
**return** $T(X)$

Fig. 8: Simulator $\mathcal{S}^{\mathbf{VRO}}$ for indifferentiability of $\mathsf{Sponge}\text{-}\mathsf{F}^{\mathbf{\Pi},\mathsf{pd}}$ (Theorem 6): variables and the interface for forward permutation query.

**public procedure** $\mathrm{P}^{-1}(Y)$
**if** $Y \in T^{-1}$ **then**
  **return** $T^{-1}(Y)$
$X \xleftarrow{\$} \{0,1\}^b$, $SC \leftarrow \mathsf{left}_r(Y) \| \mathsf{right}_c(X \oplus Y)$ // $SC \in \{0,1\}^{r+c}$
**if** $X \in T$ **then**
  **abort**
$T(X) \leftarrow Y$, $T^{-1}(Y) \leftarrow X$

Fig. 9: Simulator $\mathcal{S}^{\mathbf{VRO}}$ for Theorem 6 ($\mathsf{Sponge}\text{-}\mathsf{F}^{\mathsf{P},\mathsf{pd}}$): the interface for backward permutation query.

Indifferentiability requires to establish indistinguishability of $\Sigma_{\mathrm{id}}$ and $\Sigma_{\mathrm{re}}$. In addition, the simulator has to be efficient.

In this analysis, $\mathcal{D}$ is permitted to make additional permutation queries after finishing "normal" queries but before outputting a decision bit. Concretely, $\mathcal{D}$ can make queries to P according to the procedure of $\mathsf{Sponge\text{-}F}^{\mathrm{P},\mathsf{lfp}}$ for all construction queries $\mathrm{VOLH}(M,\nu)$, i.e., $\mathcal{D}$ can obtain all input/output pairs of $\mathcal{S}^{\mathbf{VRO}}.\mathrm{P}$ needed to calculate $\mathsf{Sponge\text{-}F}^{\mathrm{P},\mathsf{lfp}}(M,\nu)$. Note that the additional queries do not reduce the advantage of $\mathcal{D}$.

The proof in this section relies on a balls-in-bin lemma from [77, App. A] presented as follows.

**Lemma 13 (Balls-in-Bin).** *Consider a set of $|\mathcal{T}| \geq 8$ bins and $D \geq 8$ balls. Fix an integer $q \leq N$ and a sequence of integers $(\ell_1, \ldots, \ell_q)$ with $1 \leq \ell_i \leq |\mathcal{T}|$ and $\sum_{i=1}^{q} \ell_i = N$. Consider the following random process: for $i = 1, \ldots, q$, a chain of $\ell_i$ balls is thrown in consecutive bins, the initial bin being chosen independently uniformly at random. Then the probability that, at the end of the process, any bin contains $L_{max}$ balls or more, is less than $1/|\mathcal{T}|$, where*

- $L_{\max} = 2\log_2 N$ *when* $N \leq |\mathcal{T}|$;
- $L_{\max} = \frac{2N \log_2 |\mathcal{T}|}{|\mathcal{T}|}$ *when* $N \geq |\mathcal{T}|$.

### G.3 Simulator Complexity

By construction of $\mathcal{S}^{\mathbf{VRO}}$, it can be seen:

- Upon each adversarial query to $\mathrm{P}(X)$, consider the two branches at lines 8–18 and at lines 19–27 (see Fig. 8). As long as $\mathcal{S}^{\mathbf{VRO}}$ did not abort at line 7, $\mathcal{S}^{\mathbf{VRO}}$ enters at most one of the two branches. This means $\mathcal{S}^{\mathbf{VRO}}$ makes at most 1 query to **VRO** and completes at most 1 outer path.
- Upon each adversarial query to $\mathrm{P}^{-1}(Y)$, $\mathcal{S}^{\mathbf{VRO}}$ does not query **VRO** nor complete outer paths.
- $|T| \leq Q$, since $\mathcal{S}^{\mathbf{VRO}}$ defines at most 1 entry per adversarial query. Moreover, $\mathcal{S}^{\mathbf{VRO}}$ makes at most $Q$ queries to **VRO** and runs in time $O(Q)$.

Since $\mathcal{D}$ makes all internal P-queries for its construction queries (as assumed in Sect. G.2), the total number of queries to P and $\mathrm{P}^{-1}$ is the total oracle query cost metric $Q \leq p + \sigma$ introduced at the beginning of this section. This establishes the second claim of Theorem 6: $\mathcal{S}^{\mathbf{VRO}}$ makes at most $Q$ queries to **VRO**.

### G.4 Probability of Simulator Abortion

**Bad events in simulations.** We introduce the following events that may occur in a $\Sigma_{\mathrm{id}}$ execution.

- BadRW: before $\mathcal{S}^{\mathbf{VRO}}$ is to define a new rightward table entry $T(X) = Y$, any of the following is fulfilled:
  - (i) There exists another table entry $T(X') = Y'$ such that $Y = Y'$;
  - (ii) $\mathsf{right}_c(X \oplus Y) \in \{IV, IV \oplus \theta, \mathsf{right}_c(X), \mathsf{right}_c(X) \oplus \theta\}$;
  - (iii) There exists another table entry $T(X') = Y'$ such that

$$\mathsf{right}_c(X \oplus Y) \in \{\mathsf{right}_c(X'), \mathsf{right}_c(X') \oplus \theta, \mathsf{right}_c(X' \oplus Y'), \mathsf{right}_c(X' \oplus Y') \oplus \theta\}.$$

- BadLW: before $\mathcal{S}^{\mathbf{VRO}}$ is to define a new leftward table entry $T(X) = Y$, any of the following is fulfilled:
  - (i) There exists another table entry $T(X') = Y'$ such that $X = X'$;
  - (ii) $\mathsf{right}_c(X) \in \{IV, IV \oplus \theta\}$;
  - (iii) $\mathsf{right}_c(X \oplus Y) \in \{IV, IV \oplus \theta\}$;
  - (iv) There exists another table entry $T(X') = Y'$ such that

$$\mathsf{right}_c(X) \in \{\mathsf{right}_c(X' \oplus Y'), \mathsf{right}_c(X' \oplus Y') \oplus \theta\};$$

  - (v) There exists another table entry $T(X') = Y'$ such that

$$\mathsf{right}_c(X \oplus Y) \in \{\mathsf{right}_c(X'), \mathsf{right}_c(X') \oplus \theta, \mathsf{right}_c(X' \oplus Y'), \mathsf{right}_c(X' \oplus Y') \oplus \theta\}.$$

- BadAD: before $\mathcal{S}^{\mathbf{VRO}}$ is to define a new adapted table entry $T(X) \leftarrow Y$, any of the following is fulfilled:
  - (i) $Y \in T^{-1}$;
  - (ii) $(0^r \| \mathsf{right}_c(X)) \oplus Y \in T$, or $(0^r \| \mathsf{right}_c(X)) \oplus Y = X$;
  - (iii) $\mathsf{right}_c(X \oplus Y) \in \{IV, IV \oplus \theta, \mathsf{right}_c(X), \mathsf{right}_c(X) \oplus \theta\}$;
  - (iv) There exists another table entry $T(X') = Y'$ such that

$$\mathsf{right}_c(X \oplus Y) \in \{\mathsf{right}_c(X'), \mathsf{right}_c(X') \oplus \theta, \mathsf{right}_c(X' \oplus Y'), \mathsf{right}_c(X' \oplus Y') \oplus \theta\}.$$

*Probabilities of BadRW and BadLW.* For each to-be-defined rightward table entry $T(X) = Y$, the $Y$ value is newly sampled from $\{0,1\}^b$. By this, it is easy to see

$$\Pr\big[\mathsf{BadRW}\big] \le Q\Big(\frac{Q}{2^b} + \frac{4}{2^c} + \frac{4Q}{2^c}\Big) \le \frac{Q^2}{2^b} + \frac{4Q+4Q^2}{2^c}. \tag{68}$$

Similarly by symmetry, for leftward table entries it is easy to see

$$\Pr\big[\mathsf{BadLW}\big] \le Q\Big(\frac{Q}{2^b} + \frac{2}{2^c} + \frac{2}{2^c} + \frac{2Q}{2^c} + \frac{4Q}{2^c}\Big) \le \frac{Q^2}{2^b} + \frac{4Q+6Q^2}{2^c}. \tag{69}$$

*Probability of BadAD (i) and (ii).* To derive a fine-grained bound, we introduce another event ROMCol regarding adapted entries. Formally, ROMCol occurs, if there exist $\mu = 2\log_2 Q + \frac{2r'Q}{2^{r'}}$ distinct adapted entries $T(X_1) = Y_1, ..., T(X_\mu) = Y_\mu$ that collide on their rightmost $r'$ bits: $\mathsf{right}_{r'}(Y_1) = ... = \mathsf{right}_{r'}(Y_\mu)$.

To bound ROMCol, observe that by construction, every adapted table entry $T(X) = Y$ is associated with a detected path $(\overline{M}^*, SC^*, j)$ and a newly added path $(\overline{M}, SC, j+1)$: in the branch from line 8 to line 18 (in Fig. 8), the two paths can be found at lines 8 and 18 respectively; in the other branch in Fig. 8, the two paths can be found at lines 19 and 27 respectively. For convenience, we introduce an additional $b$-bit value $SC^\circ$, which is defined as

$$SC^\circ := \begin{cases} SC^* \oplus (0^r \| \theta), & \text{if } j = 0, \\ SC^*, & \text{if } j \ge 1 \end{cases} \tag{70}$$

By construction, the rightmost $r'$ bits of $Y$ is from $SC^\circ$ and $\mathbf{VRO}\big(\mathsf{unpd}(\overline{M})\big)[j+1]$:

$$\mathsf{right}_{r'}(Y) = \mathsf{right}_{r'}(SC^\circ) \oplus \mathbf{VRO}\big(\mathsf{unpd}(\overline{M})\big)[j+1].$$

By this, we associate the triple $(SC^\circ, \overline{M}, j)$ to this adapted entry.

With the above, it can be seen that the occurrence of ROMCol indicates the existence of $\mu$ "associated triples" $(SC^\circ_1, \overline{M}_1, j_1), ..., (SC^\circ_\mu, \overline{M}_\mu, j_\mu)$ that collide:

$$\begin{aligned} &\mathsf{right}_{r'}(SC^\circ_1) \oplus \mathbf{VRO}\big(\mathsf{unpd}(\overline{M}_1), r'\lambda_{\max}\big)[j_1] \\ &= ... = \mathsf{right}_{r'}(SC^\circ_\mu) \oplus \mathbf{VRO}\big(\mathsf{unpd}(\overline{M}_\mu), r'\lambda_{\max}\big)[j_\mu]. \end{aligned}$$

To bound this event, we show that for any such triple $(SC^\circ, \overline{M}, j)$, the value

$$\mathsf{right}_{r'}(SC^\circ) \oplus \mathbf{VRO}\big(\mathsf{unpd}(\overline{M}), r'\lambda_{\max}\big)[j]$$

is essentially a uniformly random string of $r'$ bits. We distinguish three cases:

- Case 1: $j \ge 2$. In this case, it essentially holds

$$\begin{aligned} &\mathsf{right}_{r'}(SC) \oplus \mathbf{VRO}\big(\mathsf{unpd}(\overline{M}), \nu_{\max}\big)[j] \\ &= \mathbf{VRO}\big(\mathsf{unpd}(\overline{M}), \nu_{\max}\big)[j-1] \oplus \mathbf{VRO}\big(\mathsf{unpd}(\overline{M}), \nu_{\max}\big)[j] \end{aligned}$$

  by the construction of $\mathcal{S}^{\mathbf{VRO}}$.
- Case 2: $j = 1$. In this case, the value $\mathsf{right}_{r'}(SC^*)$ is independent from the others $\mathbf{VRO}\big(\mathsf{unpd}(\overline{M}), \nu_{\max}\big)[2], ..., \mathbf{VRO}\big(\mathsf{unpd}(\overline{M}), \nu_{\max}\big)$

By the above and since we assumed $r' \ge 3$ and $Q \ge 8$, we can apply the balls-in-bin result Lemma 13 to bound $\Pr\big[\mathsf{ROMCol}\big]$. Concretely, by Lemma 13, $\mu = 2\log_2 Q + \frac{2r'Q}{2^{r'}}$ provides a general upper bound on the bin-size for any $Q$. Therefore,

$$\Pr\big[\mathsf{ROMCol}\big] \le \frac{1}{2^{r'}}. \tag{71}$$

We then bound BadAD (i) and (ii) conditioned on ¬ROMCol. Consider a to-be-defined adapted entry $T(X) \leftarrow Y$, and let $y_1 = \mathsf{left}_{b-r'}(Y)$ and $y_2 = \mathsf{right}_{r'}(Y)$. By construction of $\mathcal{S}^{\mathbf{VRO}}$, $y_1$ is newly sampled from $\{0,1\}^{b-r'}$. Let $\alpha_{y_2}$ be the number of table entries $T(X') = Y'$ with $\mathsf{right}_{r'}(Y) = y_2$. The probability to have $Y \in T^{-1}$, i.e., $Y = Y'$ for some $T(X') = Y'$, is then bounded by $\alpha_{y_2}/2^{b-r'}$. Furthermore, it holds $\sum_{y_2 \in \{0,1\}^{r'}} \alpha_{y_2} \le Q$.

As argued above, each adapted entry is uniquely associated with a triple $(SC^{\circ}, \overline{M}, j)$. Taking a union bound yields

$$\begin{aligned}
&\leq \Pr\big[\exists \text{ a to-be-defined adapted entry } T(X) \leftarrow Y : Y \in T^{-1} \mid \neg\mathsf{ROMCol}\big] \\
&= \sum_{(SC^{\circ},\overline{M},j)} \frac{\alpha_{\mathsf{right}_{r'}(SC^{\circ})\oplus\mathbf{VRO}(\mathsf{unpd}(\overline{M}),b\lambda_{\max})[j]}}{2^{b-r'}} \\
&= \sum_{y_2\in\{0,1\}^{r'}} \sum_{(SC^{\circ},\overline{M},j):\mathsf{right}_{r'}(SC^{\circ})\oplus\mathbf{VRO}(\mathsf{unpd}(\overline{M}),b\lambda_{\max})[j]=y_2} \frac{\alpha_{y_2}}{2^{b-r'}} \\
&\leq \mu \sum_{y_2\in\{0,1\}^{r'}} \frac{\alpha_{y_2}}{2^{b-r'}} \\
&\leq \Big(2\log_2 Q + \frac{2r'Q}{2^{r'}}\Big)\frac{Q}{2^{b-r'}} = \frac{2Q\log_2 Q}{2^{b-r'}} + \frac{2r'Q^2}{2^b}. \qquad (72)
\end{aligned}$$

By similar ideas, it can be shown that

$$\begin{aligned}
&\Pr\big[\exists \text{ an adapted entry } T(X) = Y : (0^r\|\mathsf{right}_c(X))\oplus Y \in T \mid \neg\mathsf{ROMCol}\big] \\
&\leq \frac{2Q\log_2 Q}{2^{b-r'}} + \frac{2r'Q^2}{2^b}. \qquad (73)
\end{aligned}$$

Furthermore, using the uniformness of $y_1 = \mathsf{left}_{b-r'}(Y)$, it can be shown

$$\Pr\big[\exists \text{ an adapted entry } T(X) = Y : (0^r\|\mathsf{right}_c(X))\oplus Y = X\big] \leq \frac{Q}{2^{b-r'}}. \qquad (74)$$

*Probability of BadAD (iii) and (iv), and conclusion.* By construction, right before $\mathcal{S}^{\mathbf{VRO}}$ is to define an adapted entry $T(X) = Y$, it holds: there exists a corresponding random oracle query $M$ and an index $j$, and $\mathcal{S}^{\mathbf{VRO}}$ has just sampled a random string $y_1 \xleftarrow{\$} \{0,1\}^{b-r'}$, and $Y = (0^r\|\mathsf{right}_c(X))\oplus\big(y_1\|\mathbf{VRO}(M, b\lambda_{\max})[j]\big)$. By this, it holds

$$\Pr\big[\mathsf{right}_c(X\oplus Y) = IV\big] = \frac{1}{2^c},$$

The same bound holds for the other seven types of collisions in (iii) and (iv). Therefore,

$$\Pr\big[\mathsf{BadAD} \text{ (iii), (iv)}\big] \leq Q\Big(\frac{4}{2^c} + \frac{4Q}{2^c}\Big),$$

which plus Eqs. (71), (72), (73) and (74) yield

$$\begin{aligned}
\Pr\big[\mathsf{BadAD}\big] &\leq \Pr\big[\mathsf{ROMCol}\big] + \Pr\big[\mathsf{BadAD} \mid \mathsf{ROMCol}\big] \\
&\leq \frac{1}{2^{r'}} + 2\Big(\frac{2Q\log_2 Q}{2^{b-r'}} + \frac{2r'Q^2}{2^b}\Big) + \frac{Q}{2^{b-r'}} + Q\Big(\frac{4}{2^c} + \frac{4Q}{2^c}\Big) \\
&\leq \frac{1}{2^{r'}} + \frac{5Q\log_2 Q}{2^{b-r'}} + \frac{4r'Q^2}{2^b} + \frac{4Q+4Q^2}{2^c}. \qquad (75)
\end{aligned}$$

For simplicity, define $\mathsf{Bad} := \mathsf{BadRW} \vee \mathsf{BadLW} \vee \mathsf{BadAD}$. Gathering Eqs. (68), (69) and (75) yields the final bound:

$$\Pr\big[\mathsf{Bad}\big] \leq \frac{1}{2^{r'}} + \frac{(4r'+2)Q^2}{2^b} + \frac{12Q+14Q^2}{2^c} + \frac{5Q\log_2 Q}{2^{b-r'}}. \qquad (76)$$

**Properties of good $\Sigma_{\mathbf{id}}$ executions.** We first introduce some terminology.

A *simulator cycle* consists of the execution period starting from when an adversary makes a query to when the adversary receives an answer. Moreover:

- If the simulator cycle was triggered by $\mathcal{D}$ querying $\mathrm{P}(x)$ and $\mathcal{S}^{\mathbf{VRO}}$ executes line 33 (Fig. 8) to define a table entry, then it is called a *rightward (simulator) cycle*. Table entries defined in such cycles are *rightward entries*;
- If the simulator cycle was triggered by $\mathcal{D}$ querying $\mathrm{P}^{-1}(Y)$ and $\mathcal{S}^{\mathbf{VRO}}$ executes line 7 (Fig. 9) to define a table entry, then it is called a *leftward (simulator) cycle*. Table entries defined in such cycles are *leftward entries*;
- If the simulator cycle was triggered by $\mathcal{D}$ querying $\mathrm{P}(x)$ and $\mathcal{S}^{\mathbf{VRO}}$ executes either line 17 or line 26 (Fig. 9) to define a table entry, then it is called an *adapted (simulator) cycle*. Table entries defined in such cycles are *adapted entries*.

Given an inner path $(\overline{M}, SC, 0) \in \mathit{Paths}$, $\overline{M} = \overline{M}[1]\|...\|\overline{M}[\ell]$, we say that $T(X_1) = Y_1, ..., T(X_\ell) = Y_\ell$ *are the table entries underlying the path* $(\overline{M}, SC, 0)$, if and only if:

– $\mathsf{right}_c(X_1) = IV$, $\mathsf{left}_r(X_1) = \overline{M}[1]$;
– $\mathsf{left}_r(Y_{i-1} \oplus X_i) = \overline{M}[i]$ for $i = 2, 3, ...\ell$;
– $\mathsf{right}_c(X_i \oplus Y_i) = \mathsf{right}_c(X_{i+1})$ for $i = 1, 2, ...\ell - 1$.

Given an outer path $(\overline{M}, SC, j) \in Paths$, $\overline{M} = \overline{M}[1] \| ... \| \overline{M}[\ell - j + 1]$, $1 \leq j \leq \ell$, we say that $T(X_1) = Y_1, ..., T(X_\ell) = Y_\ell$ *are the table entries underlying the path* $(\overline{M}, SC, j)$, if and only if:

– $\mathsf{right}_c(X_1) = IV$ (when $j < \ell$) or $IV \oplus \theta$ (when $j = \ell$), $\mathsf{left}_r(X_1) = \overline{M}[1]$;
– $\mathsf{left}_r(Y_{i-1} \oplus X_i) = \overline{M}[i]$ for $i = 2, 3, ...\ell - j + 1$;
– $\mathsf{right}_c(X_{i+1}) = \mathsf{right}_c(X_i \oplus Y_i)$ for all $i = 1, ..., \ell - j - 1, \ell - j + 1, ..., \ell - 1$, and $\mathsf{right}_c(X_{\ell-j+1}) = \mathsf{right}_c(X_{\ell-j} \oplus Y_{\ell-j}) \oplus \theta$.

**Lemma 14.** *In a good $\Sigma_{\mathrm{id}}$ execution, there never exist distinct paths $(\overline{M}_1, SC_1, j_1)$ and $(\overline{M}_2, SC_2, j_2)$ in Paths such that $\mathsf{right}_c(SC_1) = \mathsf{right}_c(SC_2)$ or $\mathsf{right}_c(SC_1) = \mathsf{right}_c(SC_2) \oplus \theta$. This also means $\mathrm{P}(X)$ never abort at line 7 (Fig. 8).*

*Proof.* Let $T(X_1) = Y_1, ..., T(X_{\ell_1}) = Y_{\ell_1}$ be the table entries underlying the path $(\overline{M}_1, SC_1, j_1)$, and $T(X'_1) = Y'_1, ..., T(X'_{\ell_2}) = Y'_{\ell_2}$ be the table entries underlying $(\overline{M}_2, SC_2, j_2)$, as shown in Fig. 10. Wlog assume $\ell_1 \leq \ell_2$. We exclude the possibility of each case as follows.

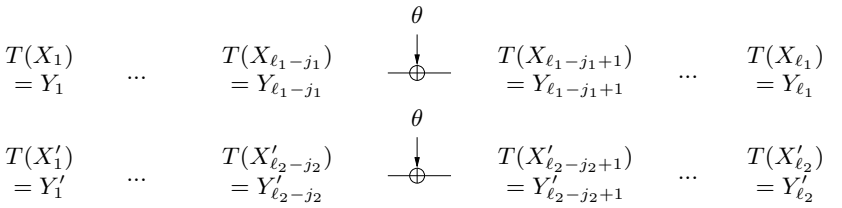


Fig. 10: Lemma 14: table entries underlying $(\overline{M}_1, SC_1, j_1)$ and $(\overline{M}_2, SC_2, j_2)$.

*Case 1: $\exists 0 \leq i \leq \ell_1 - 1$ such that $X_{\ell_1-i} \neq X'_{\ell_2-i}$.* Let $i^\circ$ be the smallest integer such that $X_{\ell_1-i^\circ} \neq X'_{\ell_2-i^\circ}$, which means $X_{\ell_1-i^\circ+1} = X'_{\ell_2-i^\circ+1}$. Then it can be seen $\mathsf{right}_c(X_{\ell_1-i^\circ} \oplus Y_{\ell_1-i^\circ}) = \mathsf{right}_c(X'_{\ell_2-i^\circ} \oplus Y'_{\ell_2-i^\circ}) = \mathsf{right}_c(X_{\ell_1-i^\circ+1})$, which indicates the occurrence of at least one of the events BadRW, BadLW and BadAD and contradicts the goodness of the execution.

*Case 2: $X_{\ell_1-i} = X'_{\ell_2-i}$ for all $0 \leq i \leq \ell_1 - 1$, but $\ell_1 \neq \ell_2$ .* When $j_2 \neq \ell_1$ it holds $\mathsf{right}_c(X'_{\ell_2-\ell_1} \oplus Y'_{\ell_2-\ell_1}) = \mathsf{right}_c(X_{\ell_1}) = IV$ (when $j_2 \neq \ell_1$; when $j_2 = \ell_1$ it holds $\mathsf{right}_c(X'_{\ell_2-\ell_1} \oplus Y'_{\ell_2-\ell_1}) = \mathsf{right}_c(X_{\ell_1}) \oplus \theta = IV \oplus \theta$, which indicates the occurrence of at least one of BadRW, BadLW and BadAD and contradicts goodness of the execution.

*Case 3: $X_{\ell_1-i} = X'_{\ell_2-i}$ for all $0 \leq i \leq \ell_1 - 1$ and $\ell_1 = \ell_2$ .* Since $(\overline{M}_1, SC_1, j_1)$ and $(\overline{M}_2, SC_2, j_2)$ are distinct, it has to be $j_1 \neq j_2$. Wlog assume $j_1 < j_2$. Then, it holds:

– $\mathsf{right}_c(X_{\ell_1-j_1}) = \mathsf{right}_c(X_{\ell_1-j_1-1} \oplus Y_{\ell_1-j_1-1}) \oplus \theta$, and
– $\mathsf{right}_c(X'_{\ell_1-j_1}) = \mathsf{right}_c(X_{\ell_1-j_1-1} \oplus Y_{\ell_1-j_1-1})$,

meaning that $\mathsf{right}_c(X'_{\ell_1-j_1}) \neq \mathsf{right}_c(X_{\ell_1-j_1})$ and contradicts the condition $X_{\ell_1-i} = X'_{\ell_2-i}$ for all $0 \leq i \leq \ell_1 - 1$. This case is thus not possible either. □

**Lemma 15.** *In a good $\Sigma_{\mathrm{id}}$ execution, all the table entries underlying every inner path are rightward.*

*Proof.* The claim holds before the first simulator cycle, since no path has $\ell \geq 1$ at that time. Now, assume that the claim holds before the $i$-th simulator cycle, and we prove that it still holds after the $i$-th cycle. We distinguish three cases.

*Case 1: $i$-th cycle is rightward.* Assume that the table entry defined in this cycle is $T(X) = Y$. This entry may "extend" existing inner paths. Though, if the claims does not hold after $\mathcal{S}^{\mathbf{VRO}}$ defining $T(X) = Y$, then there necessarily exists another (either leftward or adapted) entry $T(X') = Y'$ such that $\mathsf{right}_c(X \oplus Y) = \mathsf{right}_c(X')$. This indicates BadRW and contradicts goodness of the execution.

*Case 2: $i$-th cycle is leftward.* Assume that the entry defined in this cycle is $T(X) = Y$. Conditioned on ¬BadLW, this entry can never be the first entry of any path. By this, if $T(X) = Y$ "extends" existing inner path (and changes their state), then there necessarily exists another entry $T(X') = Y'$ such that $\mathsf{right}_c(X' \oplus Y') = \mathsf{right}_c(X)$. This indicates the occurence of Bad.

*Case 3: $i$-th cycle is adapted.* Assume that the entry defined in this cycle is $T(X) = Y$. By construction of $\mathcal{S}^{\mathbf{VRO}}$, this means right before the $i$-th cycle, there exists a path $(\overline{M}, SC, j) \in Paths$ with $\mathsf{right}_c(SC) = \mathsf{right}_c(X)$ (when $j = 0$) or $\mathsf{right}_c(SC)\oplus\theta = \mathsf{right}_c(X)$ (when $j \geq 1$). If $T(X) = Y$ "extends" another inner path $(\overline{M}', SC', 0) \in Paths$, then it has $\mathsf{right}_c(SC') = \mathsf{right}_c(X) = \mathsf{right}_c(SC)$ (when $j = 0$) or $\mathsf{right}_c(SC') = \mathsf{right}_c(SC)\oplus\theta$ (when $j \geq 1$). This contradicts Lemma 14.

*Summary.* By the above, as long as the execution is good, the claim still holds after the $i$-th simulator cycle. It thus holds throughout the execution. □

**Lemma 16.** *In a good $\Sigma_{\mathrm{id}}$ execution, consider any path $(\overline{M}, SC, j) \in Paths$, and let $T(X_1) = Y_1, ..., T(X_\ell) = Y_\ell$ be the table entries underlying $(\overline{M}, SC, j)$. Then, for all $i \in \{1, ..., \ell - 1\}$, the entry $T(X_{i+1}) = Y_{i+1}$ is defined later than the entry $T(X_i) = Y_i$.*

*Proof.* Let $M = \mathsf{unpd}(\overline{M})$. Assume otherwise, and assume that $i \in \{1, ..., \ell-1\}$ is the smallest index such that $T(X_i) = Y_i$ is defined no later than $T(X_{i+1}) = Y_{i+1}$.

*Case 1: $X_i \neq X_{i+1}$ and $T(X_i) = Y_i$ is defined later.* $\mathsf{right}_c(X_i\oplus Y_i) = \mathsf{right}_c(X_{i+1})$ or $\mathsf{right}_c(X_i\oplus Y_i) = \mathsf{right}_c(X_{i+1})\oplus\theta$ indicate the occurrence of BadRW (iii) (when $T(X_i) = Y_i$ is rightward), or BadLW (v) (when $T(X_i) = Y_i$ is leftward), or BadAD (iv) (when $T(X_i) = Y_i$ is adapted), and contradicts the goodness of the execution.

*Case 2: $X_i = X_{i+1}$.* This means $T(X_i) = Y_i$ has $\mathsf{right}_c(X_i\oplus Y_i) = \mathsf{right}_c(X_i)$ or $\mathsf{right}_c(X_i\oplus Y_i) = \mathsf{right}_c(X_i)\oplus\theta$. $T(X_i) = Y_i$ must be a leftward entry: otherwise, this indicates the occurrence of BadRW (ii) or BadAD (iii) and contradicts goodness of the execution. However, a leftward entry $T(X_i) = Y_i$ appearing in a path indicates either of the following:

- Subcase 2.1: $\mathsf{right}_c(X_i) = IV$ or $IV\oplus\theta$. This indicates BadLW (ii);
- Subcase 2.2: there exists another entry $T(X') = Y'$ with $\mathsf{right}_c(X_i) = \mathsf{right}_c(X'\oplus Y')$. This indicates BadLW (iv).

In all, Case 2 is not possible either. These complete the proof. □

**Lemma 17.** *In a good $\Sigma_{\mathrm{id}}$ execution, the simulator never aborts.*

*Proof.* Lemma 14 has shown that $\mathcal{S}^{\mathbf{VRO}}$ never aborts at line 7 in Fig. 8. The other abortions are due to collisions in to-be-defined new table entries, the occurrence of which indicates BadRW, BadRW or BadAD and contradicts goodness of the execution. □

### G.5 Consistency of Simulation

**Lemma 18.** *At the end of a good $\Sigma_{\mathrm{id}}$ execution, it holds: for every* $\mathbf{VRO}$ *query $\mathbf{VRO}(M, \nu) \to Z[1]\|...\|Z[\lambda]$ ($\lambda = \lceil \frac{\nu}{r'} \rceil$) appeared during the execution, $\overline{M}[1]\|...\|\overline{M}[\ell] = \mathsf{pd}(M)$, there exists a unique outer path $(\overline{M}[1]\|...\|\overline{M}[\ell], SC, \lambda) \in Paths$ with underlying table entries*

$$T(X_1) = Y_1, ..., T(X_{\ell+\lambda-1}) = Y_{\ell+\lambda-1},$$

*s.t. $\mathsf{right}_{|Z[i]|}(X_{\ell+i}\oplus Y_{\ell+i}) = Z[i]$ for all $i \in \{0, ..., \lambda - 1\}$. I.e., $\mathsf{Sponge}\text{-}\mathsf{F}^{T,\mathsf{pd}}(M, \nu) = \mathbf{VRO}(M, \nu)$.*

*Proof.* Depending on who made the query $\mathbf{VRO}(M, \nu) \to Z[1]\|...\|Z[\lambda]$, we distinguish two cases:

**Case 1: $\mathbf{VRO}(M, \nu) \to Z[1]\|...\|Z[\lambda]$ is firstly made by $\mathcal{S}^{\mathbf{VRO}}$.** Then by construction of $\mathcal{S}^{\mathbf{VRO}}$, it holds: (i) $\nu = \lambda r'$; (ii) the query $\mathbf{VRO}(M, \nu) \to Z[1]\|...\|Z[\lambda]$ is necessarily made in an adapted simulator cycle, in which $\mathcal{S}^{\mathbf{VRO}}$ detects a path $(\overline{M}^*, SC^*, \lambda-1)$ and tries to complete it to some path $(\overline{M}, SC, \lambda)$, $\overline{M} = \mathsf{pd}(M)$. Assume that this adapted cycle was due to $\mathcal{D}$ querying $\mathrm{P}(X)$. We distinguish two subcases:

- Subcase 1.1: $\lambda \geq 2$. Then $\overline{M} = \overline{M}^*$ by construction of $\mathcal{S}^{\mathbf{VRO}}$. Meanwhile, the last table entry underlying $(\overline{M}, SC^*, \lambda-1)$ is necessarily an adapted entry defined in an earlier adapted simulator cycle. By this, conditioned on ¬BadAD, right after $(\overline{M}, SC^*, \lambda - 1)$ is added to $Paths$, no entry $T(X') = Y'$ with $\mathsf{right}_c(X') = \mathsf{right}_c(SC^*)$ ever exists, which also means $X = \star\|\mathsf{right}_c(SC^*) \notin T$.
  From the point $(\overline{M}, SC^*, \lambda - 1)$ is added to $Paths$, at any point until $\mathcal{D}$ queries $\mathrm{P}(X)$, it is not possible that $T(X)$ is defined:
  - $T(X)$ cannot be defined due to $\mathcal{D}$ querying $\mathrm{P}(X)$, since this implies that the path $(\overline{M}, SC, \lambda)$ has been completed and the purported current adapted simulator cycle would not happen;

- $T(X)$ cannot be defined due to $\mathcal{D}$ querying $\mathrm{P}^{-1}(Y')$ for some other $Y'$, since this implies the occurrence of $\mathsf{BadLW}$ and contradicts the goodness of the $\Sigma_{\mathrm{id}}$ execution.

On the other hand, the value $Y \leftarrow (0^r \| \mathsf{right}_c(SC^*)) \oplus (y_1 \| Z[\lambda])$ derived from the random oracle response $Z[\lambda]$ and the sampled string $y_1$ has $Y \notin T^{-1}$ due to $\neg\mathsf{BadAD}$. By these, $\mathcal{S}^{\mathbf{VRO}}$ would succeed in defining an adapted entry $T(X) = Y$ that is consistent with $\mathbf{VRO}(M, \nu) \rightarrow Z[1] \| ... \| Z[\lambda]$.

– Subcase 1.2: $\lambda = 1$ and $\overline{M}^* \neq$ empty_string. Then by Lemma 16, the detected (inner) path $(\overline{M}^*, SC^*, 0)$ is fully constituted by rightward table entries; by Lemma 16, the last entry underlying $(\overline{M}^*, SC^*, 0)$ is the latest created. By these, conditioned on $\neg\mathsf{BadRW}$, right after $(\overline{M}^*, SC^*, 0)$ is added to $Paths$, no entry $T(X') = Y'$ with $\mathsf{right}_c(X') = \mathsf{right}_c(SC^*) \oplus \theta$ ever exists, which means $X = \star \| \mathsf{right}_c(SC^*) \notin T$.
Similarly to Subcase 1.1, from the point $(\overline{M}, SC^*, 0)$ is added to $Paths$, at any point until $\mathcal{D}$ queries $\mathrm{P}(X)$, it is not possible that $T(X)$ is defined. By these, conditioned on $\neg\mathsf{BadAD}$, $\mathcal{S}^{\mathbf{VRO}}$ would succeed in defining an adapted entry $T(X) = Y$ that is consistent with $\mathbf{VRO}(M, r') \rightarrow Z[1]$.

– Subcase 1.3: $\lambda = 1$ and $\overline{M}^* =$ empty_string. This means the detected path is $(\text{empty\_string}, 0^r \| IV, 0)$, and the adversarial query is $\mathrm{P}(\mathsf{pd}(M) \| IV \oplus \theta)$. At the very beginning of the execution, it certainly holds $(\mathsf{pd}(M) \| IV \oplus \theta) \notin T$. Similarly to Subcase 1.1, at any point until $\mathcal{D}$ queries $\mathrm{P}(\mathsf{pd}(M) \| IV \oplus \theta)$, it is not possible that $T(\mathsf{pd}(M) \| IV \oplus \theta)$ is defined. By these, conditioned on $\neg\mathsf{BadAD}$, $\mathcal{S}^{\mathbf{VRO}}$ will succeed in defining adapted entry $T(\mathsf{pd}(M) \| IV \oplus \theta) = Y$ that is consistent with $\mathbf{VRO}(M, r') \rightarrow Z[1]$.

Moreover, by construction, the table entries will never be changed in the subsequent execution. Therefore, in any subcase, the claim holds.

**Case 2: $\mathbf{VRO}(M, \nu) \rightarrow Z[1] \| ... \| Z[\lambda]$ is firstly made by $\mathcal{D}$.** Then, since $\mathcal{D}$ makes queries to P according to the procedure of $\mathsf{Sponge}\text{-}\mathsf{F}^{\mathrm{P},\mathsf{pd}}(M, \nu)$ for $\mathbf{VRO}(M, \nu)$ (as assumed in Sect. G.2), the path corresponding to $\mathbf{VRO}(M, \nu) \rightarrow Z[1] \| ... \| Z[\lambda]$ will eventually appear in $T$, i.e., it eventually holds $(\mathsf{pd}(M), SC, \lambda) \in Paths$. Let $T(X_1) = Y_1, ..., T(X_\ell) = Y_\ell$ be the entries underlying the path $(\overline{M}, SC, \lambda)$, $\overline{M} = \mathsf{pd}(M)$. By Lemma 16, the entries are defined from $T(X_1) = Y_1$ to $T(X_\ell) = Y_\ell$ one-by-one. None of them can be defined in leftward simulator cycles. Therefore, when $\mathcal{D}$ queried $\mathrm{P}(X_{\ell-\lambda+1})$, $\mathcal{S}^{\mathbf{VRO}}$ would have detected the path $\left(\mathsf{left}_{(\ell-\lambda)r}(\overline{M}), SC^*, 0\right)$ with $\mathsf{right}_c(SC^*) \oplus \theta = \mathsf{right}_c(X_{\ell-\lambda+1})$ and defined the adapted entry $T(X_{\ell-\lambda+1}) = Y_{\ell-\lambda+1}$ consistently with $Z[1] = \mathbf{VRO}(M, r')$ (as long as $\mathsf{BadAD}$ does not occur). By iterating this argument, it can be seen that the subsequent adapted entries $T(X_{\ell-\lambda+2}) = Y_{\ell-\lambda+2}, ..., T(X_\ell) = Y_\ell$ have $Z[2] = \mathbf{VRO}(M, 2r')[2], ..., Z[\lambda] = \mathbf{VRO}(M, \lambda r')[\lambda]$ respectively. The claim thus also holds in this case. □

### G.6 Indistinguishability of $\Sigma_{\mathrm{id}}$ and $\Sigma_{\mathrm{re}}$

We now prove indistinguishability of $\Sigma_{\mathrm{id}}$ and $\Sigma_{\mathrm{re}}$ using the randomness mapping argument [32]. For this, we make the internal randomness of the simulator $\mathcal{S}^{\mathbf{VRO}}$ explicit, by letting it access a tape $\psi$ of uniformly distributed random bits: every time $\mathcal{S}^{\mathbf{VRO}}$ is to sample $m$ random bits, it submits $m$ to $\psi$ to have the $m$ bits. As argued in [32,3], using explicit randomness is equivalent with lazy sampling, since the former can be simulated by the latter.

The underlying randomness of $\Sigma_{\mathrm{id}}$ is then fully determined by the random pair $(\mathbf{VRO}, \psi)$. Such a pair $(\mathbf{VRO}, \psi)$ is *bad*, if the event $\mathsf{Bad}$ (defined in Sect. G.4) occurs during the $\Sigma_{\mathrm{id}}$ execution $\mathcal{D}^{\Sigma_{\mathrm{id}}(\mathbf{VRO},\psi)}$ that uses $(\mathbf{VRO}, \psi)$ as the internal randomness. We define a map $\Lambda$ mapping $(\mathbf{VRO}, \psi)$ either to the special symbol $\perp$ when $(\mathbf{VRO}, \psi)$ is bad, or to a *partial permutation* defined by the simulator table $T$ when $(\mathbf{VRO}, \psi)$ is good.

Then map $\Lambda$ is defined for good random pairs $(\mathbf{VRO}, \psi)$ as follows: run $\mathcal{D}^{\Sigma_{\mathrm{id}}(\mathbf{VRO},\psi)}$, and consider the table $T$ of the simulator at the end of the execution: then fill all undefined entries of $T$ with the special symbol $*$. The result is exactly $\Lambda(\mathbf{VRO}, \psi)$. By our simulator construction, $\Lambda(\mathbf{VRO}, \psi)$ is a partial permutation. We say that a partial permutation table $T$ is good, if it has a good preimage by $\Lambda$. Then, we say that a permutation $\mathbf{\Pi}$ extends a partial permutation table $T$, denoted $\mathbf{\Pi} \vdash T$, if $\mathbf{\Pi}(X) = T(X)$ for all $X$ such that $T(X) \neq *$.

By definition of the randomness mapping, for any good partial permutation table $T$, the outputs of $\mathcal{D}^{\Sigma_{\mathrm{id}}(\mathbf{VRO},\psi)}$ and $\mathcal{D}^{\Sigma_{\mathrm{re}}(\mathbf{\Pi})}$ are equal for any $(\mathbf{VRO}, \psi)$ such that $\Lambda(\mathbf{VRO}, \psi) = T$ and any permutation $\mathbf{\Pi} \vdash T$. Let $\Omega_1$ be the set of partial permutation tables $T$ such that $\mathcal{D}^{\Sigma_{\mathrm{id}}(\mathbf{VRO},\psi)}$ output 1 for any pair $(\mathbf{VRO}, \psi)$ such that $\Lambda(\mathbf{VRO}, \psi) = T$. Then, Lemma 19 states that $\Sigma_{\mathrm{id}}$ and $\Sigma_{\mathrm{re}}$ executions linked by the same good table $T$ yield the same adversarial output, while Lemma 20 states that $\Sigma_{\mathrm{id}}$ and $\Sigma_{\mathrm{re}}$ executions linked by the same good table $T$ have close probabilities of occurrence.

**Lemma 19.** *Consider a fixed distinguisher $\mathcal{D}$, let $T$ be a good table w.r.t. $\mathcal{D}$. Then, for any random pair $(\mathbf{VRO}, \psi)$ such that $\Lambda(\mathbf{VRO}, \psi) = T$ and any random permutation $\mathbf{\Pi}$ such that $\mathbf{\Pi} \vdash T$, $\mathcal{D}$ obtains the same transcript of queries and responses in the two executions $\mathcal{D}^{\Sigma_{\mathrm{id}}(\mathbf{VRO},\psi)}$ and $\mathcal{D}^{\Sigma_{\mathrm{re}}(\mathbf{\Pi})}$, and gives the same output.*

*Proof.* Imagine running the two executions $\mathcal{D}^{\Sigma_{\mathrm{id}}(\mathbf{VRO},\psi)}$ and $\mathcal{D}^{\Sigma_{\mathrm{re}}(\mathbf{\Pi})}$ in parallel. We prove the claim via an induction on the sequence of queries of $\mathcal{D}$. Assume that the sequences of queries and answers of $\mathcal{D}$ are the same up to some point in the executions $\mathcal{D}^{\Sigma_{\mathrm{id}}(\mathbf{VRO},\psi)}$ and $\mathcal{D}^{\Sigma_{\mathrm{re}}(\mathbf{\Pi})}$, and consider the next query. Since $\mathcal{D}$ is deterministic, it issues the same query as the next. We distinguish four cases.

– Case 1: $\mathcal{D}$'s next query is to $\mathrm{P}(X)$. Then, in $\mathcal{D}^{\Sigma_{\mathrm{id}}(\mathbf{VRO},\psi)}$ the answer $Y$ will be written into the table entry $T(X)$, no matter how $Y$ is defined. In $\mathcal{D}^{\Sigma_{\mathrm{re}}(\mathbf{\Pi})}$ the answer is given by the random permutation $\mathbf{\Pi}(X)$, which equals $T(X)$ since $\mathbf{\Pi} \vdash T$.
– Case 2: $\mathcal{D}$'s next query is to $\mathrm{P}^{-1}(Y)$. This is similar to Case 1 by symmetry.
– Case 3: $\mathcal{D}$'s next query is to $\mathrm{VOLH}(M,\nu)$. In $\mathcal{D}^{\Sigma_{\mathrm{id}}(\mathbf{VRO},\psi)}$ the answer is given by $\mathbf{VRO}(M,\nu)$. By Lemma 18, it holds $\mathsf{Sponge\text{-}F}^{T,\mathsf{lfp}}(M,\nu) = \mathbf{VRO}(M,\nu)$. Since $\mathbf{\Pi} \vdash T$, the real world answer has $\mathsf{Sponge\text{-}F}^{\mathbf{\Pi},\mathsf{pd}}(M,\nu) = \mathsf{Sponge\text{-}F}^{T,\mathsf{lfp}}(M,\nu)$, and thus further equals the ideal world response $\mathbf{VRO}(M,\nu)$

In any case, the next answers are the same in the two executions. Therefore, $\mathcal{D}$ gains the same transcript of queries and answers in $\mathcal{D}^{\Sigma_{\mathrm{id}}(\mathbf{VRO},\psi)}$ and $\mathcal{D}^{\Sigma_{\mathrm{re}}(\mathbf{\Pi})}$. Since $\mathcal{D}$ is deterministic, its output is determined by this transcript, and is thus the same in the two executions. □

**Lemma 20.** *Consider a fixed distinguisher $\mathcal{D}$ with total oracle query cost at most $Q$. Then, for any $T \in \Omega_1$, it holds*

$$\frac{\Pr\big[\mathbf{\Pi} \vdash T\big]}{\Pr\big[\Lambda(\mathbf{VRO},\psi) = T\big]} \geq 1.$$

*Proof.* (i) Let $|T|$ be the number of "non-empty" entries in $T$; (ii) In $\Sigma_{\mathrm{id}}$, every entry in $T$ is defined using $b$ uniformly distributed bits obtained from $\psi$ and $\mathbf{VRO}$, and thus $\Pr\big[\Lambda(\mathbf{VRO},\psi) = T\big] = \frac{1}{2^{b|T|}}$; (iii) $\Pr\big[\mathbf{\Pi} \vdash T\big] = \frac{1}{(2^b)_{|T|}}$. Therefore,

$$\frac{\Pr\big[\mathbf{\Pi} \vdash T\big]}{\Pr\big[\Lambda(\mathbf{VRO},\psi) = T\big]} \geq \frac{2^{b|T|}}{(2^b)_{|T|}} \geq 1,$$

as claimed. □

**Lemma 21.** *For any distinguisher $D$ with total oracle query cost at most $Q$, it holds*

$$\Big|\Pr\big[\mathcal{D}^{\Sigma_{\mathrm{id}}} = 1\big] - \Pr\big[\mathcal{D}^{\Sigma_{\mathrm{re}}} = 1\big]\Big| \leq \frac{1}{2^{r'}} + \frac{(4r'+2)Q^2}{2^b} + \frac{12Q+14Q^2}{2^c} + \frac{5Q\log_2 Q}{2^{b-r'}}.$$

*Proof.* Gathering Lemmas 19, 20 and Eq. (76) yields

$$\begin{aligned}
&\Big|\Pr\big[\mathcal{D}^{\Sigma_{\mathrm{id}}} = 1\big] - \Pr\big[\mathcal{D}^{\Sigma_{\mathrm{re}}} = 1\big]\Big| \\
\leq &\Pr\big[(\mathbf{VRO},\psi) \textit{ is bad}\big] + \sum_{T\in\Omega_1} \Pr\big[\Lambda(\mathbf{VRO},\psi) = T\big] - \sum_{T\in\Omega_1} \Pr\big[\mathbf{\Pi} \vdash T\big] \\
\leq &\Pr\big[(\mathbf{VRO},\psi) \textit{ is bad}\big] \\
&\quad + \sum_{T\in\Omega_1} \Pr\big[\Lambda(\mathbf{VRO},\psi) = T\big]\left(1 - \frac{\Pr\big[\mathbf{\Pi} \vdash T\big]}{\Pr\big[\Lambda(\mathbf{VRO},\psi) = T\big]}\right) \\
\leq &\frac{1}{2^{r'}} + \frac{(4r'+2)Q^2}{2^b} + \frac{12Q+14Q^2}{2^c} + \frac{5Q\log_2 Q}{2^{b-r'}} + \sum_{T\in\Omega_1} \Pr\big[\Lambda(\mathbf{VRO},\psi) = T\big] \\
\leq &\frac{1}{2^{r'}} + \frac{(4r'+2)Q^2}{2^b} + \frac{12Q+14Q^2}{2^c} + \frac{5Q\log_2 Q}{2^{b-r'}},
\end{aligned}$$

as claimed. □

# H Instances and Performances

## H.1 Ascon-p-based Instances and Hardware Performances

Denote by Ascon-p the (12-round) permutation used in the NIST lightweight standard Ascon. Sun et al. [88] proposed two Ascon-p-based instances, namely, Ascon-DM and Ascon-DM-128: the former aims at 192-bit 2nd preimage security for challenge messages up to $2^{64}$ bits, while the latter aims at 128-bit 2nd preimage security for challenge messages up to $2^{64}$ bits. In an earlier version [87], Sun et al. proposed two other Ascon-p-based instances Ascon-EDM and Ascon-EDM-128 for 192-bit and 128-bit 2nd preimage security respectively. As mentioned in Table 1 caption, Ascon-EDM and Ascon-EDM-128 are explicitly stated as not suggested [88, Sect. 1.4].

To enable a fair comparison with Sun et al. [88], we also propose two instances of $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$ using Ascon-p: (i) The instance Ascon-Sp-F sets $\mathsf{P}$ = Ascon-p and parameters $c = r' = h = 256$ and $r = 64$; (ii) The faster instance Ascon-SpFwd-f sets $\mathsf{P}$ = Ascon-p and parameters $c = r' = h = 128$ and $r = 192$. The parameters and provable security bounds of these instances and Ascon-Hash256 are summarized in Table 2.

**Table 2.** Comparison between Ascon-Sp-F, Ascon-SpFwd-f; Ascon-DM, Ascon-DM-128, Ascon-EDM, Ascon-EDM-128 of Sun et al. [88] and Ascon-Hash256. All parameters are expressed in bits. There is no quantum security proof for Ascon-DM. For (quantum) second preimage security (q-)spre., security is considered for challenge messages with up to $2^{64}$ blocks (after padding).

| Algorithm | State | $h$ | $r$ | $c$ | Col. | Pre. | spre. | q-col. | q-pre. | q-spre. | Ref. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Ascon-Hash256 | 320 | 256 | 64 | 256 | 128 | 192 | 128 | 26.7 | 26.7 | 26.7 | [84] |
| Ascon-DM | 640 | 256 | 64 | 256 | 128 | 192 | 192 | - | - | - | [88] |
| Ascon-DM-128 | 640 | 128 | 128 | 192 | 64 | 128 | 128 | - | - | - | [88] |
| Ascon-EDM | 576 | 256 | 64 | 256 | 128 | 192 | 192 | - | - | - | [87] |
| Ascon-EDM-128 | 576 | 128 | 128 | 192 | 64 | 128 | 128 | - | - | - | [87] |
| Ascon-Sp-F | 576 | 256 | 64 | 256 | 128 | 256 | 192 | 64 | 128 | 96 | Ours |
| Ascon-SpFwd-f | 576 | 128 | 128 | 192 | 64 | 128 | 128 | 32 | 64 | 64 | Ours |
| SHA-3-512 | 1600 | 512 | 576 | 1024 | 256 | 512 | 512 | 133.3 | 133.3 | 133.3 | [84] |
| Keccak-DM[576, 512] | 3200 | 512 | 1024 | 576 | 256 | 512 | 512 | - | - | - | [88] |
| Keccak-EDM$^c$[576, 512] | 2176 | 512 | 1024 | 576 | 256 | 512 | 512 | - | - | - | [88] |
| Keccak-Sp-F$_{512}$ | 2176 | 512 | 1024 | 576 | 256 | 512 | 512 | 128 | 256 | 256 | Ours |
| Keccak-DM[832, 768] | 3200 | 768 | 768 | 832 | 384 | 768 | 768 | - | - | - | [88] |
| Keccak-EDM$^c$[832, 768] | 2432 | 768 | 768 | 832 | 384 | 768 | 768 | - | - | - | [88] |
| Keccak-Sp-F$_{768}$ | 2432 | 768 | 768 | 832 | 384 | 768 | 768 | 192 | 384 | 384 | Ours |
| Keccak-DM[1088, 1024] | 3200 | 1024 | 512 | 1088 | 512 | 1024 | 1024 | - | - | - | [88] |
| Keccak-EDM$^c$[1088, 1024] | 2688 | 1024 | 512 | 1088 | 512 | 1024 | 1024 | - | - | - | [88] |
| Keccak-Sp-F$_{1024}$ | 2688 | 1024 | 512 | 1088 | 512 | 1024 | 1024 | 256 | 512 | 512 | Ours |

**Hardware performances of Ascon-Hash256, Ascon-EDM and Ascon-Sp-F.** To compare the hardware performance, we implemented three of the algorithms Ascon-Hash256, Ascon-EDM and Ascon-Sp-F in Verilog HDL using a round-based architecture for permutation. We give the performance of hash implementations by using Synopsis Design Compiler with NanGate 45nm and TSMC 90nm standard cell libraries. For a more accurate and fair comparison, we gave both area-optimized and delay-optimized implementations. The hardware performance results are summarized in Table 3. To evaluate throughput, we tested the implementations with different message lengths, with results shown in Table 4. Compared with Ascon-Hash256, despite that the added feed-forward increases hardware area, Ascon-Sp-F achieves better delay and throughput for short messages (due to its single-call squeezing). Concretely, for a message with $\ell$ blocks (after padded), Ascon-Sp-F consumes $2 + 12\ell$ cycles to process, whereas Ascon-Hash256 and Ascon-EDM need $2+12(\ell+3)$ cycles. Consequently, for messages with less than 64 bytes (most of the hash computations in the LMS signature are indeed processing less than 64 bytes: see App. I.1), Ascon-Sp-F achieves a far better throughput.

**Table 3.** Hardware performance of different hash implementations, including area- and delay-optimized.

| Hash | Area-Optimized | | | Delay-Optimized | | |
|---|---|---|---|---|---|---|
| | Area[GE] | Delay[ns] | Power[uW] | Area[GE] | Delay[ns] | Power[uW] |
| Ascon-Hash256 | 8141.00 | 1.35 | 64.5050 | 9114.33 | 0.39 | 18.7290 |
| Ascon-EDM | 11571.67 | 1.51 | 90.1044 | 13519.67 | 0.48 | 20.7705 |
| Ascon-Sp-F | 12512.33 | 1.66 | 96.2426 | 14922.67 | 0.51 | 21.11 |

### H.2 Keccak-p-based Instances

Denote by Keccak-p the (24-round) permutation used in SHA-3. Sun et al. [88] proposed a series of Keccak-p-based instances of their constructions Sponge-DM$^{\mathsf{P},\mathsf{pd}}$ and Sponge-EDM$^{\mathsf{P},\mathsf{pd}}$ for different levels of preimage security for challenge messages up to $2^{64}$ bits, including 224-bit, 256-bit, 384-bit, 512-bit, 768-bit and 1024-bit security. For each security level, we could define analogue instances of Sponge-F$^{\mathsf{P},\mathsf{pd}}$. For simplicity and fair comparison, we focus on achieving 512-bit, 768-bit and 1024-bit preimage security for challenge messages up to $2^{64}$ bits, and propose three instances of Sponge-F$^{\mathsf{P},\mathsf{pd}}$ using Keccak-p: (i) The least secure instance Keccak-Sp-F$_{512}$ sets $\mathsf{P}$ = Keccak-p and parameters $r' = h = 512$, $c = h + 64 = 576$ and $r = 1024$; (ii) The medium instance Keccak-Sp-F$_{768}$ sets $\mathsf{P}$ = Keccak-p

**Table 4.** Throughput of different hash implementations, including area- and delay-optimized. For the sake of space, Ascon-Hash256, Ascon-EDM and Ascon-Sp-F are abbreviated as A-Hash, A-EDM and A-SpFwd respectively.

| MsgLen (Bytes) | Area-Optimized | | | Delay-Optimized | | |
|---|---|---|---|---|---|---|
| | A-Hash | A-EDM | A-SpFwd | A-Hash | A-EDM | A-SpFwd |
| 4 | 382.3178017 | 423.8410596 | 1376.936317 | 1323.407775 | 1333.333333 | 4481.792717 |
| 8 | 640.6406406 | 683.6146123 | 1482.854495 | 2217.602218 | 2150.537634 | 4826.546003 |
| 12 | 960.960961 | 1025.421918 | 2224.281742 | 3326.403326 | 3225.806452 | 7239.819005 |
| 16 | 1102.497847 | 1145.516377 | 2029.169309 | 3816.3387 | 3603.603604 | 6604.747162 |
| 28 | 1693.121693 | 1724.934545 | 2698.795181 | 5860.805861 | 5426.356589 | 8784.313725 |
| 64 | 2400.375059 | 2322.416765 | 2803.943045 | 8308.990587 | 7305.936073 | 9126.559715 |
| 128 | 2986.29338 | 2802.25494 | 2994.502281 | 10337.16939 | 8815.426997 | 9746.811346 |

and parameters $r' = h = 768$, $c = h + 64 = 832$ and $r = 768$; (iii) The most secure instance Keccak-Sp-F$_{1024}$ sets $\mathsf{P}$ = Keccak-p and parameters $r' = h = 1024$, $c = h + 64 = 1088$ and $r = 512$. Keccak-Sp-F$_{512}$ has the same level of collision and (second) preimage security as SHA-3-512 for messages with at most 64 blocks, but is expected to be 1.77 times that of SHA-3-512 due to its larger rate. Keccak-Sp-F$_{768}$ and Keccak-Sp-F$_{1024}$ could meet the requirements of 768-bit and 1024-bit (second) preimage security in Chinese call for new hash [73]. The parameters and provable security bounds of these instances and SHA-3-512 are summarized in Table 2 as well.

# I LMS Signature

## I.1 Description of LMS Signature

LMS is a two level signature scheme, where a one time signature (LM-OTS) is used to sign the message, while a Merkle tree signs the LM-OTS public key.

**Description of LM-OTS.** We follow the description given in [39, Sect. 2.2]. We begin with a detailed description of the LM-OTS scheme. Let $\mathsf{h} : \{0,1\}^* \to \{0,1\}^h$. Fix $w \in \{1,2,4,8\}$ as a parameter of the scheme, and set $e \leftarrow 2^w - 1$. Set $u \leftarrow s/w$; note that the output of $\mathsf{h}$ can be a sequence of $u$ integers, each $w$ bits long. Set $v := \lceil \lfloor \log u \cdot e + 1 \rfloor / w \rceil$ and $p := u + v$. Define a function checksum : $(\{0,1\}^w)^u \to \{0,1\}^{wv}$ as follows:

$$\text{checksum}(d_0, \ldots, d_{u-1}) := \sum_{i=0}^{u-1} (e - d_i)$$

where each $d_i \in \{0,1\}^w$ is viewed as an integer in the range $\{0, \ldots, 2^w - 1\}$ and the result is expressed as an integer using exactly $wv$ bits. For positive integers $i, m$ with $i < 2^{8m}$, we let $[i]_m$ denote the $m$-byte representation of $i$ in bigendian order. For a string $s$ and positive integer $j$, set $H^0_{I,q}(x; j) := x$. For positive integers $i \geq 1$ and $j$, define

$$H^i_{I,q,d}(x; j) := \mathsf{h}\big(I \parallel [q]_4 \parallel [d]_2 \parallel [i + j - 1]_1 \parallel H^{i-1}_{I,q,d}(x; j)\big)$$

Define the LM-OTS scheme as follows:

*Key-generation algorithm GenOTS.* Key Generation takes as input $I, q$, where $I$ is a 16 byte identifier, and $q$ is a 4 byte diversification factor. The algorithm proceeds as follows:

1. Choose $p$ uniform values $x_0, \ldots, x_{p-1} \in \{0,1\}^s$.
2. For $i = 0$ to $p - 1$, compute $y_i = H^e_{I,q,i}(x_i; 0)$.
3. Compute $pk := \mathsf{h}(I \parallel [q]_4 \parallel [8080]_2 \parallel y_0 \parallel \ \ldots \ \parallel y_{p-1})$

The public key is $pk$, and the private key is $sk = (x_0, \ldots, x_{p-1})$.

*Signing algorithm SignOTS.* Signing inputs a private key $sk = (x_0, \ldots, x_{p-1})$ and a message $M \in \{0,1\}^*$ as usual, as well as $I, q$ as above. It does:

1. Choose uniform $C \in \{0,1\}^s$.
2. Compute $Q := \mathsf{h}(I \parallel [q]_4 \parallel [8181]_2 \parallel C \parallel M)$ and $c := \text{Checksum}(Q)$. Set $V := Q \| c$, and parse $V$ as a sequence of $w$-bit integers $V_0, \ldots, V_{p-1}$
3. For $i = 0, \ldots, p - 1$, compute $\sigma_i := H^{V_i}_{I,q,i}(x_i; 0)$
4. Return the signature $\sigma = (C, q, \sigma_0, \ldots, \sigma_{p-1})$

*Verification algorithm VrfyOTS.* Verification takes as input a message $M \in \{0,1\}^*$ and a signature $(C, q, \sigma_0, \ldots, \sigma_{p-1})$, as well as $I$ and $q$ as above. It does:

1. Compute $Q := \mathsf{h}(I \parallel [q]_4 \parallel [8181]_2 \parallel C \parallel M)$ and $c := \text{Checksum}(Q)$ Set $V := Q \| c$, and parse $V$ as a sequence of $w$-bit integers $V_0, \ldots, V_{p-1}$
2. For $i = 0, \ldots, p-1$, compute $\sigma_i := H^{V_i}_{I,q,i}(\sigma_i; V_i)$
3. Output $\mathsf{h}(I \parallel [q]_4 \parallel [(8080)]_2 \parallel y_0 \parallel \ldots \parallel y_{p-1})$

We note that, in contrast to the usual convention, VrfyOTS returns a string rather than a bit and does not take a public key as input. A signature $\sigma$ on some message $M$ is valid relative to some fixed public key $pk$ if the output of VrfyOTS is equal to $pk$.

One can verify that correctness holds in the following sense: for any $I, q$, and $(sk, pk)$ output by $\text{GenOTS}(I, q)$, and any message $M$, we have:

$$\text{VrfyOTS}(\text{SignOTS}(sk, M, I, q), I) = pk.$$

**Description of LMS.** An instance of the LMS scheme is defined by computing a Merkle tree of height $h$ using $2^h$ LM-OTS public keys at the leaves. We give a formal definition now.

Let $H : \{0,1\}^* \to \{0,1\}^s$ as previously. We fix $w \in \{1, 2, 4, 8\}$ for use in the LM-OTS system, and we also select an integer $h$.

*Key-generation algorithm Gen.* Key Generation takes as input a parameter $d$ and a value $I \in \{0,1\}^{128}$. The algorithm proceeds as follows:

1. For $q = 0, \ldots, 2^d - 1$, compute $(pk^q, sk^q) \leftarrow \text{GenOTS}(I, q)$.
2. For $r = 2^d, \ldots, 2^{d+1} - 1$, set $T[r] := \mathsf{h}(I \parallel [r]_4 \parallel [8282]_2 \parallel pk^{r-2^d})$
3. For $r = 2^d - 1, \ldots, 1$, set $T[r] := \mathsf{h}(I \parallel [r]_4 \parallel [8383]_2 \parallel T[2r] \parallel T[2r+1])$

The public key is $pk = (d, I, T[1])$, and the private key is

$$sk = \left(sk^0, \ldots, sk^{2^d-1}, T[1..2^{d+1} - 1]\right).$$

*Signing algorithm Sign* takes a private key $\left(sk^0, \ldots, sk^{2^d-1}, T[1..2^{d+1} - 1]\right)$, an integer $0 \le q < 2^d$, a message $M \in \{0,1\}^*$, as well as $I$ as above as inputs. The algorithm proceeds as follows:

1. Compute $\sigma := \text{SignOTS}(sk^q, M, I, q)$.
2. Set $p_0, \ldots, p_{d-1}$ as $p_i := T[\lfloor (q + 2^d)/2^i \rfloor \oplus 1]$
3. Return the signature $\Sigma = (\sigma, p_0, \ldots, p_{d-1})$

*Verification algorithm Vrfy.* Verification takes as input a public key $(d, I, T)$, a message $M \in \{0,1\}^*$, and a signature $\Sigma = (\sigma, p_0, \ldots, p_{d-1})$ and an integer $0 \le q < 2^d$. The algorithm proceeds as follows:

1. Compute $pk := \text{VrfyOTS}(M, \sigma)$.
2. Set $r := q + 2^d$, and compute $T[r] = \mathsf{h}(I, [r]_4, [8282]_2, pk)$
3. For $i = 1..d$, set $r := \lfloor (q + 2^d)/2^i \rfloor$, and compute $T[r]$ as follows:
   - $T[r] := \mathsf{h}(I \parallel [r]_4 \parallel [8383]_2 \parallel T[2r] \parallel p_{i-1})$ if $\lfloor q/2^{i-1} \rfloor$ is even
   - $T[r] := \mathsf{h}(I \parallel [r]_4 \parallel [8383]_2 \parallel p_{i-1} \parallel T[2r+1])$ if $\lfloor q/2^{i-1} \rfloor$ is odd
4. Return 1 if and only if $T[1] = T$.

This verification procedure works because, if the signature is valid, all $T$ elements computed during the verification procedure match the corresponding $T$ elements of the private key.

### I.2 Remark on Existing Provable Security Results of LMS

Modeling the hash function as a monolithic (quantum) variable-input-length (VIL) random oracle, Katz [53] and Eaton [38] proved classical and quantum security for LMS respectively. SHAKE is sufficiently close to such a random oracle, in the sense that its sponge structure is indifferentiable from a VIL random oracle. However, SHA-256 structure is *not* indifferentiable from a VIL random oracle [31], and Katz [53] and Eaton [38] results could not imply SHA-256-based LMS generic security, in a strict sense.

To capture the influences of SHA-256 structure, Fluhrer modeled the SHA-256 compression function as a fixed-input-length (FIL) random oracle and proved generic security for SHA-256-based LMS via a dedicated proof [39, Corollary 1]. Our Theorem 5 followed this approach, by modeling Ascon-p as a random permutation and proving generic security for Ascon-Sp-F-based LMS directly.

In many signature constructions, hash functions only process *public inputs* (in particular, the signed message), and SHA-256 structure is indifferentiable from such a public-use random oracle [35] and this enables deriving security proofs for SHA-256-based signature constructions. However, in LMS (and the underlying Winternitz OTS scheme), the signature secret key is the input of some hash calls, meaning that [35] is *inapplicable*. What's worse, even if [35] is applicable, the public indifferentiability bound of SHA-256 structure is $q^2/2^{256}$, and this only indicates an upper bound of $2^{-16}$ on the attack success probability for the case $q = 2^{120}$. This is far below both the claims in [39, Corollary 1] and our dedicated result Theorem 5.

Fluhrer's proof proceeded with two steps: first, proving multi-target (second) preimage security (Definition 1) on Merkle-Damgård with injective padding [39, Theorem 1, Corollary 1], which captures the structure of SHA-256; second, deriving generic security for SHA-256-based LMS. Unfortunately, along the way, we found that Fluhrer's multi-target (second) preimage security provable results on Merkle-Damgård with injective padding [39, Theorem 1, Corollary 1] are *flawed*, due to neglecting the influence of second preimage challenge message length. For Merkle-Damgård with injective padding we exhibit a concrete attack breaking the claims in [39, Theorem 1, Corollary 1] and RFC 8554 [64, Sect. 9]. In the next paragraphs, we first serve a description of the Merkle-Damgård construction used by LMS, and then elaborate our findings in detail.

**Definition of Merkle-Damgård Construction $\mathsf{MD}^{\mathbf{F},\mathsf{apd}}$** The Merkle-Damgård construction $\mathsf{MD}^{f,pad}$ is built upon a compression function $f : \{0,1\}^h \times \{0,1\}^r \mapsto \{0,1\}^h$ and padding function $pad : \{0,1\}^* \mapsto (\{0,1\}^r)^{\ell_{\max}}$. It uses a fixed initialization vector $IV \in \{0,1\}^h$. Upon input a message $M \in \{0,1\}^*$, $\mathsf{MD}^{f,pad}(M)$ proceeds as follows:

(i) $S \leftarrow IV$
(ii) $(\overline{M}[1], \ldots, \overline{M}[\ell]) \xleftarrow{r} pad(M)$
(iii) **for** $i = 1, 2, ..., \ell$ **do**
  – $S \leftarrow f(S, M[i])$
(iv) Output $Z \leftarrow S$

**Flaw in Fluhrer's Result.** As a model of SHA-256 in LMS [64,29], Fluhrer considered a Merkle-Damgård hash construction $\mathsf{MD}^{\mathbf{F},\mathsf{apd}}$ (a definition is give in App. I.2) using a *random function* $\mathbf{F} : \{0,1\}^h \times \{0,1\}^r \mapsto \{0,1\}^h$ and an *injective appending-padding* scheme $\mathsf{apd}$ (as defined in Sect. 2), and proved the following bound.

**Theorem 7 (Theorem 1 of [39]).** *Let $\Phi$ be the following constraints:*

*(i) Any $\mathsf{prefix}_{1,i} \in \mathcal{G}_1$ and the message $M^{1,i}$ satisfy $|\mathsf{apd}(\mathsf{prefix}_{1,i}\|M^{1,i})| = r$ (which means $\mathsf{MD}^{\mathbf{F},\mathsf{apd}}(\mathsf{prefix}_{1,i}\|M^{1,i})$ only makes 1 call to $\mathbf{F}$);*
*(ii) Any $\mathsf{prefix}_{2,i} \in \mathcal{G}_2$ satisfies $|\mathsf{prefix}_{2,i}| \le r$ (which means $\mathsf{prefix}_{2,i}$ can be absorbed within 1 call to $\mathbf{F}$);*
*(iii) Any $\mathsf{prefix}_{3,i} \in \mathcal{G}_3$ satisfies $|\mathsf{prefix}_{3,i}| \le r - h$ (which means $\mathsf{prefix}_{3,i}\|R^i$ has at most $r$ bits and can be absorbed within 1 call to $\mathbf{F}$).*

*Then, when the total number of $\mathbf{F}$-queries appeared during $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{MD}^{\mathbf{F},\mathsf{apd}},\mathscr{A},\mu,\Phi}$ is $Q$, for any integer threshold $C$ it holds*

$$\mathbf{Adv}^{\text{lms-mfpspr}[\Phi]}_{\mathsf{MD}^{\mathbf{F},\mathsf{apd}}}(\mathscr{A}) \le \frac{(2\mu+1)(C+1)Q}{2^h} + \frac{1}{C!}\cdot\left(\frac{Q^2}{2^{h+1}}\right)^C. \tag{77}$$

For SHA-256 parameter $h = 256$ in the case of $\mu = 3$, Fluhrer recommended setting $C = 7$, yielding a success probability upper bound of $2^{-129}$.

Unfortunately, this bound is *incorrect*, at least for some injective appending-padding. For simplicity, consider the simplest case where: (i) $\mathsf{apd}(M) = \mathsf{pd10}^*(M)$ (as defined in Sect. 2); (ii) $\mathcal{G}_1 = \mathcal{G}_2 = \emptyset$ and $\mathcal{G}_3 = \{\mathsf{prefix}\}$, $|\mathsf{prefix}| = r - h$ (i.e., only one prefix of $r - h$ bits is available for Group-3 challenges). This means $\mu = 1$. We further assume that the attacker $\mathscr{A}^{\mathbf{F}}$ chooses a sufficiently long $M$ with $(\ell - 1)r \le |\mathsf{prefix}\|R\|M| \le \ell r - 1$, so that $\mathsf{apd}(\mathsf{prefix}\|R\|M)$ has $\ell$ $r$-bit blocks, and obtains a single Group-3 challenge $Z = \mathsf{MD}^{\mathbf{F},\mathsf{apd}}(\mathsf{prefix}\|R\|M)$. By Fig. 4 the definition of lms-mfpspr, $\mathscr{A}^{\mathbf{F}}$ wins as long as it finds $M' \neq R\|M$ with $\mathsf{MD}^{\mathbf{F},\mathsf{apd}}(\mathsf{prefix}\|M') = \mathsf{MD}^{\mathbf{F},\mathsf{apd}}(\mathsf{prefix}\|R\|M)$.

Let $(\overline{M}[1], ..., \overline{M}[\ell]) \xleftarrow{r} \mathsf{apd}(\mathsf{prefix}\|R\|M)$, $S_0 = IV$, $S_i = \mathbf{F}(S_{i-1}, \overline{M}[i])$ for $i = 1, ..., \ell$ be the $h$-bit intermediate values of $\mathsf{MD}^{\mathbf{F},\mathsf{apd}}(\mathsf{prefix}\|R\|M)$. Further recall that $|\mathsf{prefix}\|R| = r$, meaning that $\mathsf{prefix}\|R = \overline{M}[1]$. $\mathscr{A}^{\mathbf{F}}$ proceeds as follows:

1. Pick $R'$, $|R'| = |R|$, as well as $q$ message blocks $W_1, ..., W_q \in \{0,1\}^r$;
2. Query $S'_1 \leftarrow \mathbf{F}(IV, \mathsf{prefix}\|R')$, and then $S'_{2,i} \leftarrow \mathbf{F}(S'_1, W_i)$ for $i = 1, ..., q$;
4. Search for $(i, j) \in \{1, ..., q\} \times \{1, ..., \ell - 1\}$ with a collision $S'_{2,i} = S_j$. Once found, set $\overline{M} \leftarrow \mathsf{prefix}\|R'\|W_i\|\overline{M}[j + 1]\|...\|\overline{M}[\ell]$, $\mathsf{prefix}\|M' \leftarrow \mathsf{unapd}(\overline{M})$ and output $M'$.

Step 4 is feasible because: (i) for a message with more than 2 blocks, the $10^*$-padding and its unpadding does not "disturb" the first block $\mathsf{prefix}\|R'$; (ii) it can be seen that any string in $(\{0,1\}^r)^*$ ended with $\overline{M}[\ell]$ is a valid padded message w.r.t. the $10^*$-padding. The probability to have a pair $(i, j)$ with $S'_{2,i} = S_j$ is roughly $(\ell - 1)q/2^h$. This violates Eq. (77), in which no term depends on $\ell$ the number of challenge message blocks.

More concretely, for SHA-256 we have $h = 256$ and $\ell \le \ell_{\max} \le 2^{55} + 1$. Therefore, using $\ell = 2^{50}$ and $q = 2^{120}$, the success probability is roughly $2^{-86}$, which is far greater than the bound $2^{-129}$ given in [39, Corollary 1].

**The** m-unforge **experiment** $\mathbf{Exp}^{\text{m-unforge}}_{\mathbf{Hash}^{\mathbf{\Pi}},\mathscr{A}}$**:**

1. **Initialization:**
   1. Sample a random permutation $\mathbf{\Pi} \xleftarrow{\$} \mathcal{P}(b)$;
   2. For $\mathsf{idx} = 1, ..., u$, sample $I^{\mathsf{idx}} \xleftarrow{\$} \{0,1\}^{2r}$ and invoke $(pk^{\mathsf{idx}}, sk^{\mathsf{idx}}) \leftarrow \mathsf{Gen}(I^{\mathsf{idx}})$;
2. **Interaction:** $(\mathsf{idx}, j, M, \Sigma) \leftarrow \mathscr{A}^{\mathsf{muSign}^{\mathbf{\Pi}},\mathbf{\Pi},\mathbf{\Pi}^{-1}}((I^1, pk^1), ..., (I^u, pk^u))$, where the oracle $\mathsf{muSign}^{\mathbf{\Pi}}(\mathsf{idx}, j, M^{i,j})$ returns $\mathsf{Sign}^{\mathbf{\Pi}}(sk^{\mathsf{idx}}, j, M^{\mathsf{idx},j}, I^{\mathsf{idx}})$
   - **Adversarial restriction**: for each pair $(\mathsf{idx}, j)$, $\mathscr{A}^{\mathsf{muSign}^{\mathbf{\Pi}},\mathbf{\Pi},\mathbf{\Pi}^{-1}}$ makes at most one query.
3. **Finalization:** Output 1 if and only if $M \neq M^{\mathsf{idx},j} \wedge \mathsf{Vrfy}^{\mathbf{\Pi}}(pk^{\mathsf{idx}}, M, \Sigma, j) = 1$ (i.e., $\Sigma$ is a valid signature on $M$ for the $\mathsf{idx}$-th user with the $j$-th index).

**Fig. 11:** Experiment for multi-user unforgability of LMS-like signatures.

**Influence on SHA-256-based LMS.** We remark that if Merkle-Damgård strengthening is used, then our attacks become ineffective. Therefore, our attacks *have no threat* on the SHA-256-based LMS of RFC 8554 and SP 800-208.[9] However, in RFC 8554 [64, Sect. 9], it writes:

> *If we have no more than* $2^{64}$ *randomly chosen LMS private keys, allow the attacker access to a signing oracle and a SHA-256 hash compression oracle, and allow a maximum of* $2^{120}$ *hash compression computations, then the probability of an attacker being able to generate a single forgery against any of those LMS keys is less than* $2^{-129}$.

This actually cites [39, Corollary 1]. Our attack violates this claim, thus showing that RFC 8554 (as well as [39, Corollary 1]) is flawed and should be revised.

### I.3 Multi-user Security of LMS using Ascon-Sp-F

We consider the multi-user unforgability experiment defined in Fig. 11, which is a random permutation-based variant of the security models used by Katz [53, Sect. 3.2] and Fluhrer [39, Sect. 2.3].[10] W.r.t. this experiment, the security claim is as follows.

**Theorem 8.** *For any threshold* $\mu$ *and* $C$*, the probability that the adversary would succeed in creating a forgery is bounded by* $\frac{8u^2+4\mu C^2\ell_{\max}Q+2^{r+2}(\mu C+1)Q+84\mu ChQ}{2^h} + \frac{4Q^3}{2^{2h}} + \left(\frac{2Q^2}{2^h}\right)^C \cdot \frac{1}{C!} + \frac{8Q^2}{h!2^h} + \frac{4\mu C^2\ell_{\max}^2 Q}{2^b} + \frac{u^{\mu+1}}{2^{128\mu}} \cdot \frac{1}{(\mu+1)!}$.

*Proof.* To see that we can still apply Fluhrer's idea, we quick sketch his proof of [39, Theorem 2]. The multi-user unforgability experiment $\mathbf{Exp}^{\text{m-unforge}}_{\mathsf{h}^{\mathbf{\Pi}},\mathscr{A}}$ induces an instance of the lms-mfpspr experiment $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{h}^{\mathbf{\Pi}},\mathscr{A},\mu,\Phi}$:

- The actions in the **Initialization** phase of $\mathbf{Exp}^{\text{m-unforge}}_{\mathsf{h}^{\mathbf{\Pi}},\mathscr{A}}$ (Fig. 11) are "wrapped" as **Phase 1** of $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{h}^{\mathbf{\Pi}},\mathscr{A},\mu,\Phi}$ (Fig. 4). By Fig. 11, during the **Initialization** phase of $\mathbf{Exp}^{\text{m-unforge}}_{\mathsf{h}^{\mathbf{\Pi}},\mathscr{A}}$, $\mathsf{Gen}(I^{\mathsf{idx}})$ is invoked for all $\mathsf{idx} \in \{1, ..., u\}$, and $\mathsf{Gen}(I^{\mathsf{idx}})$ internally makes a number of calls to GenOTS and $\mathsf{h}$ (for constructing Merkle trees). During this process, every hash evaluation of the form $H^e_{I,q,i}(x_i; 0)$ (step I.1 in GenOTS) constitutes a Group-1 challenge, while every hash evaluation of the form $\mathsf{h}(I \parallel [q]_4 \parallel [8080]_2 \parallel y_0 \parallel \ldots \parallel y_{p-1})$ (step I.1 in GenOTS) constitutes a Group-2 challenge.[11] On the other hand:
  - When hashing the final Winternitz values together, the prefix $I^{\mathsf{idx}}\|q\|8080$ is formed, and this constitutes a Group-2 challenge using the final Winternitz values as the challenge messages;[12]
  - When hashing the Merkle tree leaves, for each leaf node $r$ it forms the prefix $I^{\mathsf{idx}}\|r\|8282$, and this constitutes a Group-2 challenge using the OTS public key as the challenge message;
  - When hashing the Merkle tree nodes, for each tree node $r$ it forms the prefix $I^{\mathsf{idx}}\|r\|8383$, and this constitutes a Group-2 challenge using the two child nodes as the challenge message.

[9] Patching the security proof for SHA-256-based LMS is beyond the scope of this paper, and is left for future work.

[10] Note that these definitions are "pre-challenge-query-free" to some extent: they invoke $\mathsf{Gen}$ *right after* the underlying ideal primitive, and the adversary cannot query the ideal primitive *before* the user keys are generated. By this, in the induced lms-mfpspr experiment $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{h}^{\mathbf{\Pi}},\mathscr{A},\mu,\Phi}$, the adversary cannot query $\mathbf{\Pi}$*before* generating challenge digests for $\mathcal{G}_1$ and $\mathcal{G}_2$ (**Phase 1**). We actually believe that such "pre-challenge-query-freeness" is *not* realistic: even if we don't consider the preprocessing attack model, the underlying primitives (e.g., Ascon-p) have been fixed by relative intentional standards—and the adversary can clearly compute them—*before* some concrete user keys are generated. However, fixing this definition-level issue is far out of the scope of this paper, and we have to leave it for future work.

[11] Fluhrer [39, Proof of Theorem 2] did not expand on this step, which actually constitutes Group-1 challenges.

[12] Fluhrer [39, Theorem 2] has a typo here, by mis-writing it as $I^{\mathsf{idx}}\|q\|8181$.

– The actions in the **Interaction** phase of $\mathbf{Exp}^{\text{m-unforge}}_{\mathsf{h}^{\mathbf{\Pi}},\mathscr{A}}$ (Fig. 11) are "wrapped" as **Phase 2** of $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{h}^{\mathbf{\Pi}},\mathscr{A},\mu,\Phi}$ (Fig. 4):
  - When the adversary submits a value to sign a message $M$ for the $\mathsf{idx}$-th user, index $q$, the prefix $I^{\mathsf{idx}}\|q\|8181$ is formed, and a Group-3 challenge along with the message $M$ is constituted.[13]
  - When the checksum of the hash and the Winternitz digits are computed, a number of Group-1 hash evaluations appear. However, these are *not* new challenges, meaning that this phase *only* generates new Group-3 challenges.

We now show that all the constraints of Theorem 5 are met using the Ascon-Sp-F hash function (so we have $b = 320, h = c = 256, r = 64$ and use the $1\|0^*$ padding function).

Note that for every prefix $\mathsf{prefix}_{i,j}$ generated as above, there exists a user index $\mathsf{idx} \in \{1, ..., u\}$ and a tail $tail \in \{0,1\}^{48} \cup \{0,1\}^{56}$ such that $\mathsf{prefix}_{i,j} = I^{\mathsf{idx}}\|tail$, where $|I^{\mathsf{idx}}| = 128 = 2r$. By this, the constraints $\Phi$ required in Theorem 5 are fulfilled.

It remains to prove that the three constraints required in Fig. 4 are fulfilled with high probability.

*Constraint* (1.a) *on frequency.* For the $\mathsf{idx}$-th user, each generated prefix $\mathsf{prefix}$ has $\mathsf{prefix} = I^{\mathsf{idx}}\|tail$ for some $tail \in \{0,1\}^{48} \cup \{0,1\}^{56}$. Moreover, the prefixes generated by the $\mathsf{idx}$-th user are distinct, i.e., if $\mathsf{prefix}_{i,j} = I^{\mathsf{idx}}\|tail$ and $\mathsf{prefix}_{i',j'} = I^{\mathsf{idx}}\|tail'$ are both generated by the $\mathsf{idx}$-th user, then it necessarily holds $tail \neq tail'$. This means if two generated prefixes $\mathsf{prefix}_{i,j} = \mathsf{prefix}_{i',j'}$ collide, then it must hold:

(i) $\mathsf{prefix}_{i,j} = I^{\mathsf{idx}}\|tail$ and $\mathsf{prefix}_{i',j'} = I^{\mathsf{idx}'}\|tail$ are generated by two distinct users, and
(ii) The two user identifiers $I^{\mathsf{idx}} = I^{\mathsf{idx}'}$ collide.

Therefore, as long as

$$\max_{II} \left|\left\{\mathsf{idx} \in \{1, ..., \mathsf{idx}\} : I^{\mathsf{idx}} = II\right\}\right| \leq \mu,$$

The constraint (1.a) (Fig. 4) on prefix frequency is fulfilled. Since $I^1, ..., I^u$ are independently and uniformly picked from $\{0,1\}^{128}$, it can be shown

$$\Pr\left[\max_{II} \left|\left\{\mathsf{idx} \in \{1, ..., \mathsf{idx}\} : I^{\mathsf{idx}} = II\right\}\right| > \mu\right] \leq \frac{u^{\mu+1}}{(\mu+1)!2^{128\mu}}, \tag{78}$$

meaning that constraint (1.a) in Fig. 4 is fulfilled with probability at least $1 - \frac{u^{\mu+1}}{(\mu+1)!2^{128\mu}}$.

*Constraint* (1.b) *on prefix-freeness.* By definition, if a prefix $\mathsf{prefix}_{i,j}$ is the proper prefix of another prefix $\mathsf{prefix}_{i',j'}$, then it must be that $|\mathsf{prefix}_{i,j}| = 176$ and $|\mathsf{prefix}_{i',j'}| = 184$. This means $\mathsf{prefix}_{i,j}$ is in either $\mathcal{G}_2$ or $\mathcal{G}_3$, while $\mathsf{prefix}_{i',j'}$ is in $\mathcal{G}_1$. This further means $\mathsf{prefix}_{i,j} = I^{\mathsf{idx}}\|r\|[a]_2$ for some $r \in \{0,1\}^{32}$ and $a \in \{8080, 8181, 8282, 8383\}$, whereas $\mathsf{prefix}_{i,j} = I^{\mathsf{idx}}\|q\|[i]_2\|v$ for some $i \in \{0, ..., p-1\}$ and $v \in \{0,1\}^8$. By the LMS parameter set [64, Table 1], it can be seen that $p \leq 265$. By this, such two prefixes necessarily have $[a]_2 \neq [i]_2$, meaning that the case where $\mathsf{prefix}_{i,j}$ is the proper prefix of $\mathsf{prefix}_{i',j'}$ cannot occur.

*Constraint* (1.c) *on disjointness.* By definition, prefixes in distinct groups have distinct tails, meaning that the constraint (1.c) in Fig. 4 is fulfilled.

*From* lms-mfpspr *to multi-user unforgability.* If the adversary generates a forgery in the experiment $\mathbf{Exp}^{\text{m-unforge}}_{\mathsf{h}^{\mathbf{\Pi}},\mathscr{A}}$, then the adversary could find a second preimage for the experiment $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}^{\text{Ascon-p},\mathsf{pd10}^*},\mathscr{A},\mu,\Phi}$. By the above analyses, in the experiment $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}^{\text{Ascon-p},\mathsf{pd10}^*},\mathscr{A}}$, inputs to the Ascon-p-based construction $\mathsf{Sponge\text{-}F}^{\text{Ascon-p},\mathsf{pd10}^*}$ satisfy the constraints required for Theorem 5, except with probability $\leq \frac{u^{\mu+1}}{(\mu+1)!2^{128\mu}}$ (Eq. (78)). Once they satisfy the constraints, the probability that the adversary finds a second preimage for $\mathbf{Exp}^{\text{lms-mfpspr}}_{\mathsf{Sponge\text{-}F}^{\text{Ascon-p},\mathsf{pd10}^*},\mathscr{A},\mu,\Phi}$ is bounded by Theorem 5. Summing over the two probabilities yields the claimed multi-user unforgability bound (for any $\mu, C$). □

**Corollary 1.** *Assume that* $\ell_{\max} \leq 2^{64}$ *(recall from Sect. 2 that* $\mathsf{sha}$ *length limit* $2^{55}+1$*). If we have no more than* $2^{64}$ *randomly chosen LMS private keys, allow the adversary access to a signing oracle and an* Ascon-p *permutation oracle, and allow a maximum of* $2^{120}$ Ascon-p *computations, then the probability of an adversary being able to generate a single forgery against any of those LMS keys is less than* $2^{-64}$*.*

*Proof.* For Ascon-Sp-F we have $r = 64 = h/4$. When $u = 2^{64}$, $Q = 2^{120}$ (as considered in [39, Corollary 1]), $\mu = 3$ and $C = 6$, the dominate term are bounded as follows:

(i) $\frac{4\mu C^2 \ell_{\max} Q}{2^h} \leq \frac{432 \times 2^{184}}{2^{256}} \leq \frac{1}{2^{63}}$;

[13] Fluhrer [39, Theorem 2] mis-wrote it as $I^{\mathsf{idx}}\|q\|8080$.

(ii) $\frac{2^{r+2}(\mu C+1)Q}{2^h} \le \frac{19\times 2^{186}}{2^{256}} \le \frac{1}{2^{65}}$;
(iii) $\left(\frac{2Q^2}{2^h}\right)^C \cdot \frac{1}{C!}$
(iv) $\frac{8Q^2}{h!2^h} \le \frac{8\times 2^{240}}{256!2^{256}} \le \frac{1}{2^{128}}$;
(v) $\frac{4\mu C^2\ell_{\max}^2 Q}{2^b} \le \frac{432\times 2^{248}}{2^{320}} \le \frac{1}{2^{63}}$;
(vi) $\frac{u^{\mu+1}}{2^{128\mu}} \cdot \frac{1}{(\mu+1)!} \le \frac{2^{256}}{2^{384}} \cdot \frac{1}{4!} \le \frac{1}{2128}$.

Therefore, the overall success probability is below $2^{-64}$. □

# J Security is Capped By $c$

**Preimage and Collision Attacks in the case $r' > c$.** This appendix exhibits collision and preimage attacks, showing that security is capped by $c$ and setting $r' > c$ does not buy more security.

*Preimage attack with complexity $2^c < 2^{r'}$.* Given a challenge image $Z \in \{0,1\}^{r'}$, the attack proceeds as follows:

(i) Randomly pick $q$ distinct $(Y_{11}, Y_{12}), ..., (Y_{q1}, Y_{q2}) \in \{0,1\}^{b-r'} \times \{0,1\}^c$ and query $\mathsf{P}^{-1}(Y_{11}\|\mathsf{left}_{r'-c}(Z)\|Y_{12}) \to X_1, ..., \mathsf{P}^{-1}(Y_{q1}\|\mathsf{left}_{r'-c}(Z)\|Y_{q2}) \to X_q$;
(ii) Search for $X_i$ such that $\mathsf{right}_c(X_i)\oplus Y_{i2} = \mathsf{right}_c(Z)$;
(iii) Once such an $X_i$ is found, randomly pick $q$ distinct $\overline{M}_1[1], ..., \overline{M}_{q'}[1] \in \{0,1\}^r$ and query $\mathsf{P}(\overline{M}_1[1]\|IV) \to W_1, ..., \mathsf{P}(\overline{M}_{q'}[1]\|IV) \to W_q$;
(iv) Search for $IV\oplus\mathsf{right}_c(W_j) = \mathsf{right}_c(X_i)$. Once such a $W_j$ is found, set $\overline{M}[2] \leftarrow \mathsf{left}_r(W_j)\oplus\mathsf{left}_r(X_i)$ and output $\mathsf{unpd}(\overline{M}[1]\|\overline{M}[2])$ as the preimage of $Z$.

With roughly $2q = 2\cdot 2^c$ queries, such a pair $(X_i, W_j)$ can be found. The attack complexity is thus $O(2^c)$. Clearly, this idea can also find second preimage with $O(2^c)$ complexity.

*Collision attack with complexity $2^{c/2} < 2^{r'/2}$.* The attack also leverages $\mathsf{P}^{-1}$, and proceeds as follows:

(i) Randomly pick $q$ distinct $\overline{M}_1[1], ..., \overline{M}_q[1] \in \{0,1\}^r$ and query $\mathsf{P}(\overline{M}_1[1]\|IV) \to W_1, ..., \mathsf{P}(\overline{M}_q[1]\|IV) \to W_q$;
(ii) Search for $W_i, W_j$ such that $\mathsf{right}_c(W_i)\oplus IV = \mathsf{right}_c(W_j)\oplus IV$;
(iii) Once such a pair $(W_i, W_j)$ is found, fix $X \in \{0,1\}^r$ in arbitrary, set $\overline{M}_i[2] \leftarrow \mathsf{left}_r(W_i)\oplus X$, $\overline{M}_j[2] \leftarrow \mathsf{left}_r(W_j)$ and outputs $M_1 = \mathsf{unpd}(\overline{M}_i[1]\|M[2])$ and $M_2 = \mathsf{unpd}(\overline{M}_j[1]\|M[2])$ as the pair of messages with collided digest.

With roughly $2^{c/2}$ queries, such a pair $(W_i, W_j)$ can be found. The attack complexity is thus $2^{c/2}$.

# K Illustration of Earlier Design Sponge-$\mathsf{P}^{\mathsf{P},\mathsf{pd}}$

# L Applying Human Ignorance and CI to SPONGE-DM$^{\mathsf{P},\mathsf{pd}}$

We also applied the Human Ignorance approach to the SPONGE-DM$^{\mathsf{P},\mathsf{pd}}$ construction of Sun et al. [88] shown in Fig. 13, with capacity $c$ and rate $r$.

## L.1 Single Block Squeezing Case

We first consider the simpler case of squeezing a single block output $Z = Z[1]$, $|Z| = r$. For convenience, define

$$\mathsf{SPONGE\text{-}DM}^{\mathsf{P},\mathsf{pd}}_r(M) := \mathsf{SPONGE\text{-}DM}^{\mathsf{P},\mathsf{pd}}(M, r).$$

Consider an variant of truncated Davies-Meyer construction that outputs the *leftmost* bits, i.e.,

$$\mathsf{TrDML}^{\mathsf{P}}_\nu(X) := \mathsf{left}_\nu\big(X\oplus\mathsf{P}(X)\big).$$

The security of SPONGE-DM$^{\mathsf{P},\mathsf{pd}}_r$ is closely related to TrDML$^{\mathsf{P}}_r$.

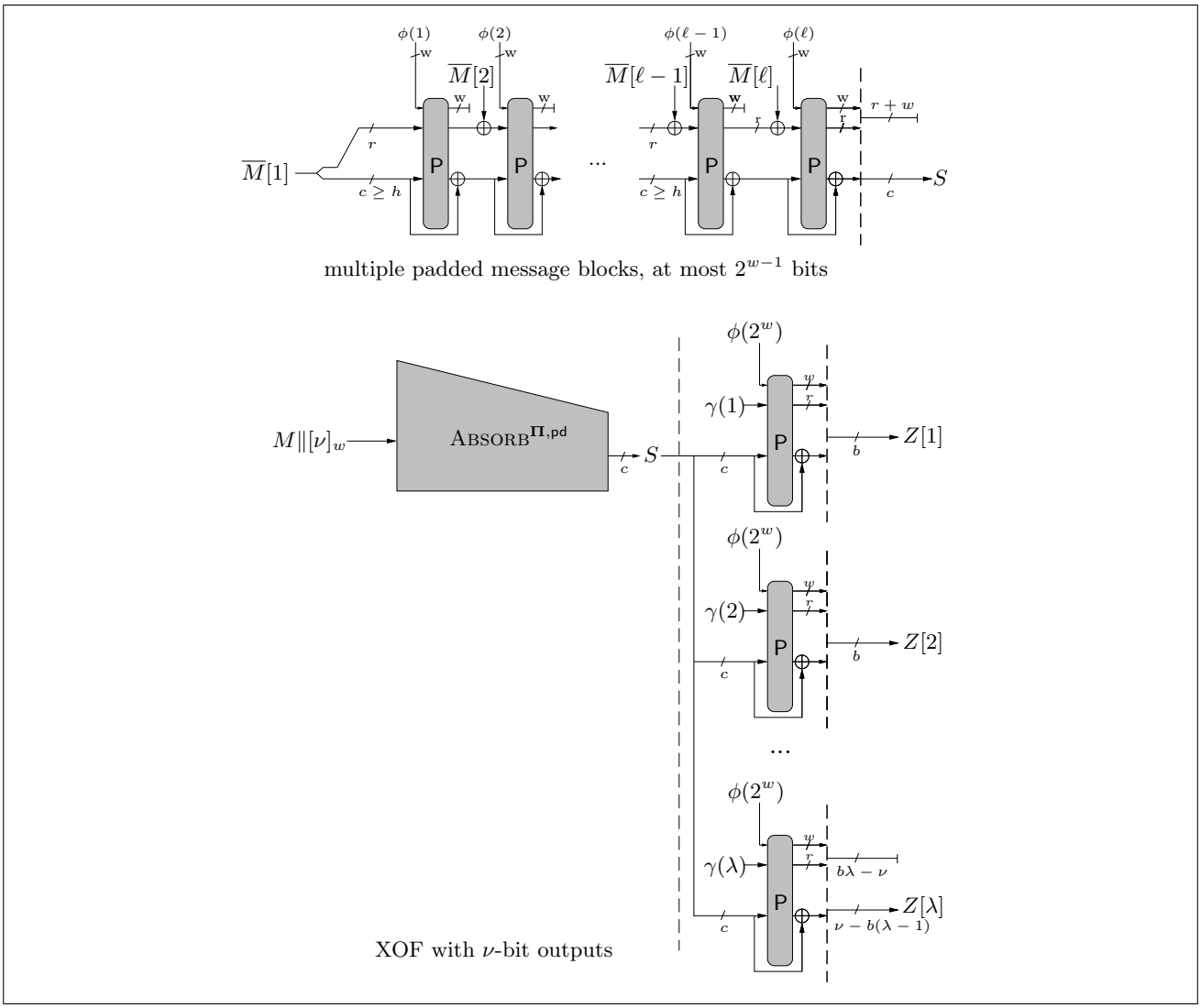


**Fig. 12:** Free-IV construction Sponge-$\mathsf{P}^{\mathsf{P},\mathsf{pd}}$ from sponge-with-feed-forward, where $\lambda = \lceil \frac{\nu}{b} \rceil$. $\phi() : \{1, ..., 2^w\} \mapsto \{0,1\}^w$ and $\gamma() : \{1, ..., 2^r\} \mapsto \{0,1\}^r$ are two injective counter-derivation functions (e.g., $\phi(i)$ can be the $w$-bit encoding of the integer $i$). (Up) Absorb function $\textsc{Absorb}^{\mathbf{\Pi},\mathsf{pd}}$; (Bottom) The full Sponge-$\mathsf{P}^{\mathsf{P},\mathsf{pd}}$ construction.

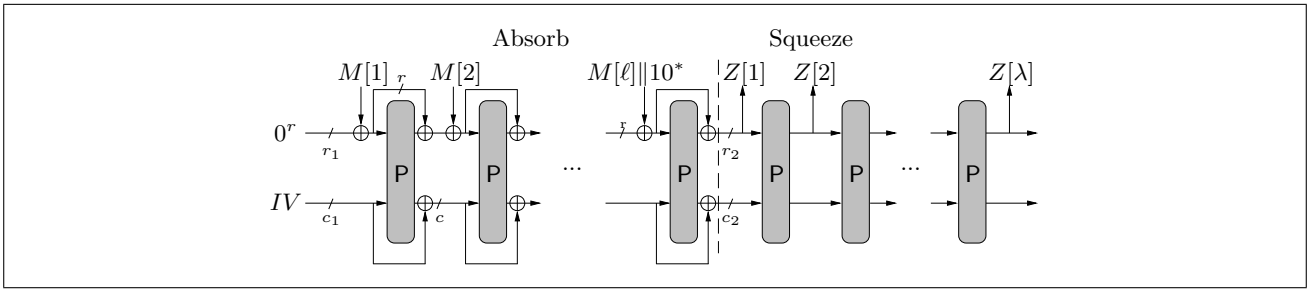


**Fig. 13:** The SPONGE-$\mathsf{DM}^{\mathsf{P},\mathsf{pd}}$ construction of Sun et al. [88]. The final output is of $\lambda r_2$ bits $Z[1]\|...\|Z[\lambda]$.

**Collision security.** We consider both the relation $R_{\mathrm{coll}}$ from Sect. 3 and a variant $R_{\mathrm{collL}}$:

$$\big((X,X'),(\mathsf{P}(X),\mathsf{P}(X'))\big) \in R_{\mathrm{collL}} \\ \textbf{iff } X \neq X' \wedge \big(\mathsf{left}_r(X{\oplus}\mathsf{P}(X)) = \mathsf{left}_r(X'{\oplus}\mathsf{P}(X'))\big). \tag{79}$$

It can be seen a collision adversary $\mathscr{A}$ outputting $M \neq M'$ with $\mathsf{SPONGE\text{-}DM}_r^{\mathsf{P},\mathsf{pd}}(M) = \mathsf{SPONGE\text{-}DM}_r^{\mathsf{P},\mathsf{pd}}(M')$ can be transformed into adversaries against $R_{\mathrm{coll}}$-ci, $R_{\mathrm{collL}}$-ci or $R_{\mathrm{pre}}(IV)$-ci.

**Lemma 22 (Collision).** *Let $(\mathsf{P},\mathsf{P}^{-1})$ be a permutation oracle (not necessarily random) and its inverse, $\mathsf{X} \in \{\mathit{classical}, \mathit{quantum}\}$, and let $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ be a $(q,t,s,\ell_{\max})$-bounded $\mathsf{X}$ collision adversary against $\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}$. Then, we can construct a $(q+2\ell_{\max}, t+O(\ell_{\max}), s+O(\ell_{\max}))$-$\mathsf{X}$ collision adversary $\mathscr{B}_{\mathrm{coll}}^{\mathsf{P},\mathsf{P}^{-1}}$ against $\mathsf{TrDM}_h^{\mathsf{P}}$ and a $(q+2\ell_{\max}, t+O(\ell_{\max}), s+O(\ell_{\max}))$-preimage adversary $\mathscr{B}_{\mathrm{pre}}^{\mathsf{P},\mathsf{P}^{-1}}$ against $\mathsf{TrDM}_c^{\mathsf{P}}$, such that:*

$$\begin{aligned}\mathbf{Adv}^{\mathrm{coll}}_{\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}}(\mathscr{A}) &\leq \mathbf{Adv}^{\mathrm{coll}}_{\mathsf{TrDM2}_r^{\mathsf{P}}}(\mathscr{B}_{\mathrm{coll}}) + \mathbf{Adv}^{\mathrm{coll}}_{\mathsf{TrDM}_h^{\mathsf{P}}}(\mathscr{B}_{\mathrm{coll}}) + \mathbf{Adv}^{IV\text{-}\mathrm{pre}}_{\mathsf{TrDM}_c^{\mathsf{P}}}(\mathscr{B}_{\mathrm{pre}}) \\ &= \mathbf{Adv}^{R_{\mathrm{collL}}\text{-}\mathrm{ci}}_{\mathsf{P}}(\mathscr{B}_{\mathrm{coll}}) + \mathbf{Adv}^{R_{\mathrm{coll}}\text{-}\mathrm{ci}}_{\mathsf{P}}(\mathscr{B}_{\mathrm{coll}}) + \mathbf{Adv}^{R_{\mathrm{pre}}(IV)\text{-}\mathrm{ci}}_{\mathsf{P}}(\mathscr{B}_{\mathrm{pre}}).\end{aligned}$$

*Proof.* Suppose that $\mathscr{A}$ outputs $(M_1,M_2)$ as a collision of $\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}$. Let $\mathsf{P}(X_1) = Y_1, ..., \mathsf{P}(X_{\ell_1}) = Y_{\ell_1}$ be the permutation-calls in evaluating $\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}(M_1)$, and let $\mathsf{P}(X'_1) = Y'_1, ..., \mathsf{P}(X'_{\ell_2}) = Y'_{\ell_2}$ be the calls in $\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}(M_2)$. If $X_{\ell_1} \neq X'_{\ell_2}$, then by outputting $(X_{\ell_1}, X'_{\ell_2})$, $\mathscr{B}_{\mathrm{coll}}$ succeeds in finding a collision for $\mathsf{TrDML}_r^{\mathsf{P}}$. If there exists $j \in \{1, ..., \min\{\ell_1,\ell_2\}-1\}$ such that $X_{\ell_1-j} \neq X'_{\ell_2-j}$, then by outputting $(X_{\ell_1-j}, X'_{\ell_2-j})$ for the smallest such $j$, $\mathscr{B}_{\mathrm{coll}}$ succeeds in finding a collision for $\mathsf{TrDM}_h^{\mathsf{P}}$. Otherwise, if $\ell_1 > \ell_2$ then $X_{\ell_1-\ell_2+1} = X'_1 = \star \| IV$, and $\mathscr{B}_{\mathrm{pre}}$ succeeds by outputting $X_{\ell_1-\ell_2}$ which equals $(\mathsf{TrDM}_c^{\mathsf{P}})^{-1}(IV)$; if $\ell_1 < \ell_2$ then $X'_{\ell_2-\ell_1+1} = X_1 = \star \| IV$, and $\mathscr{B}_{\mathrm{pre}}$ succeeds by outputting $X'_{\ell_2-\ell_1}$. This proves Eq. (20), which implies Eq. (21) by Lemma 1. Both $\mathscr{B}_{\mathrm{coll}}$ and $\mathscr{B}_{\mathrm{pre}}$ have to run $\mathscr{A}$ and then make at most $2\ell_{\max}$ $\mathsf{P}$-queries to have the above values, and the claims on their complexities follow this. □

**Preimage security (as per the definition in Sect. 2.1).** For every $Z \in \{0,1\}^r$, finding $(\mathsf{SPONGE\text{-}DM}_r^{\mathsf{P},\mathsf{pd}})^{-1}(Z)$ implies finding a preimage $(\mathsf{TrDML}_r^{\mathsf{P}})^{-1}(Z)$ for the aforementioned truncated Davies-Meyer variant, and the latter is equivalent with finding $X$ satisfying the following relation:

$$(X, \mathsf{P}(X)) \in R_{\mathrm{preL}}(Z) \textbf{ iff } \mathsf{left}_r(X{\oplus}\mathsf{P}(X)) = Z. \tag{80}$$

It can be seen that the $Z$-preimage security of $\mathsf{SPONGE\text{-}DM}_r^{\mathsf{P},\mathsf{pd}}$ can be reduced to the $R_{\mathrm{preL}}(Z)$-ci security of $\mathsf{P}$, i.e., $\mathbf{Adv}^{Z\text{-}\mathrm{pre}}_{\mathsf{SPONGE\text{-}DM}_r^{\mathsf{P},\mathsf{pd}}}(\mathscr{A}) \leq \mathbf{Adv}^{R_{\mathrm{preL}}(Z)\text{-}\mathrm{ci}}_{\mathsf{P}}(\mathscr{B})$.

**Second preimage security (as defined in Sect. 2.1).** For every $Y \in \{0,1\}^b$ we consider a $\ell_{\max}$-ary relation $R_{\text{spr-sp}}(X,Y)$:

$$(X', \mathsf{P}(X')) \in R_{\mathrm{spr}}(X,\nu) \textbf{ iff } X' \neq X,\ \mathsf{left}_\nu(X{\oplus}\mathsf{P}(X)) = \mathsf{right}_\nu(X'{\oplus}\mathsf{P}(X')). \tag{81}$$

We need to adapt the free-of-inner-collision and free-of-IV-preimage properties introduced in Sect. 4.2. Consider a second preimage challenge $M \in \{0,1\}^*$:

- $M$ is *free-of-inner-collision2*, if the permutation calls $\mathsf{P}(X_1) = Y_1, ..., \mathsf{P}(X_\ell) = Y_\ell$ underlying the evaluation $\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}(M)$ are such that:
  - $\mathsf{right}_h(X_1{\oplus}Y_1), ..., \mathsf{right}_h(X_\ell{\oplus}Y_\ell)$ are distinct, and
  - $\mathsf{left}_r(X_1{\oplus}Y_1), ..., \mathsf{right}_r(X_\ell{\oplus}Y_\ell)$ are distinct.
- $M$ is *free-of-IV-preimage*, if the permutation calls $\mathsf{P}(X_1) = Y_1, ..., \mathsf{P}(X_\ell) = Y_\ell$ underlying $\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}(M)$ are such that $IV \notin \{\mathsf{right}_c(X_1{\oplus}Y_1), ..., \mathsf{right}_c(X_\ell{\oplus}Y_\ell)\}$.

**Lemma 23 (Second Preimage).** *Let $(\mathsf{P},\mathsf{P}^{-1})$ be a permutation oracle (not necessarily random) and its inverse. We can construct a classical procedure $\textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}$ such that: for every challenge message $M \in \{0,1\}^{bl}$ that is both free-of-inner-collision and free-of-IV-preimage, and every $(q,t,s,\ell_{\max})$-bounded $\mathsf{X}$ second preimage adversary $\mathscr{A}^{\mathsf{P},\mathsf{P}^{-1}}$ against $\mathsf{Sponge\text{-}F}_h^{\mathsf{P},\mathsf{pd}}$ for the challenge $M$, $\mathsf{X} \in \{\mathit{classical}, \mathit{quantum}\}$, the procedure $\textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}(M) \Rightarrow (X_1, ..., X_\ell)$ outputs $\ell \leq \ell_{\max}$ distinct challenge inputs in $\{0,1\}^b$, and we can construct:*

*(i) A $(q+2\ell_{\max}, t_A+O(\ell_{\max}), s_A+O(\ell_{\max}))$-$\mathsf{X}$ multi-target second preimage adversary $\mathscr{B}_{\mathrm{mspr}^*}^{\mathsf{P},\mathsf{P}^{-1}}$ against $\mathsf{TrDM}^{\mathsf{P}}$ for the challenge inputs $(X_1, ..., X_{\ell-1})$;*

*(ii) A $(q+2\ell_{\max}, t_A+O(\ell_{\max}), s_A+O(\ell_{\max}))$-preimage adversary $\mathscr{B}_{\mathrm{pre}}^{\mathsf{P},\mathsf{P}^{-1}}$ against $\mathsf{TrDM}_c^{\mathsf{P}}$ for the challenge image $IV$.*

*And for* $(R_1, ..., R_{\ell-1}, R_\ell) = (R_{\rm spr}(X_1, c), ..., R_{\rm spr}(X_{\ell-1}, c), R_{\rm spr}(X_\ell, h))$*, it holds:*

$$\begin{aligned}
&\mathbf{Adv}^{M\text{-spr}}_{\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}_h}(\mathscr{A}) \\
&\le \mathbf{Adv}^{(X_\ell)\text{-mspr}^*}_{\mathsf{TrDM2}^{\mathsf{P}}_r}(\mathscr{B}_{\text{mspr}^*}) + \mathbf{Adv}^{(X_1,...,X_{\ell-1})\text{-mspr}^*}_{\mathsf{TrDM}^{\mathsf{P}}_c}(\mathscr{B}_{\text{mspr}^*}) + \mathbf{Adv}^{IV\text{-pre}}_{\mathsf{TrDM}^{\mathsf{P}}_c}(\mathscr{B}_{\text{pre}}) \\
&= \mathbf{Adv}^{R_{\rm spr}(X_\ell,r)\text{-ci}}_{\mathsf{P}}(\mathscr{B}_{\text{mspr}^*}) + \mathbf{Adv}^{(R_1,...,R_{\ell-1})\text{-mci}}_{\mathsf{P}}(\mathscr{B}_{\text{mspr}^*}) + \mathbf{Adv}^{R_{\rm pre}(IV)\text{-spr}}_{\mathsf{P}}(\mathscr{B}_{\text{pre}}).
\end{aligned}$$

*Proof.* The procedure $\textsc{InnerVals}^{\mathsf{P},\mathsf{pd}}(M)$ evaluates $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}_h(M)$ to have the underlying calls $\mathsf{P}(X_1) = Y_1, ..., \mathsf{P}(X_\ell) = Y_\ell$ and outputs $X_1, ..., X_\ell$.

Suppose that $\mathscr{A}(M)$ outputs $M'$ for $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}_h(M') = \mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}_h(M)$. Let $\mathsf{P}(X'_1) = Y'_1, ..., \mathsf{P}(X'_{\ell_2}) = Y'_{\ell_2}$ be the calls in $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}_h(M')$. Then:

- If $\mathsf{left}_r(X_{\ell_1} \oplus Y_{\ell_1}) = \mathsf{left}_r(X'_{\ell_2} \oplus Y'_{\ell_2})$, then by outputting $X'_{\ell_2}$, $\mathscr{B}_{\rm spr}$ finds a second preimage $\mathsf{TrDML}^{\mathsf{P}}_r(X'_{\ell_2}) = \mathsf{TrDML}^{\mathsf{P}}_r(X_{\ell_1})$;
- If there exists $j \in \{1, ..., \min\{\ell, \ell_2\} - 1\}$ such that $X_{\ell-j} \neq X'_{\ell_2-j}$, then by outputting $X'_{\ell_2-j}$, $\mathscr{B}_{\text{mspr}^*}$ finds a second preimage $\mathsf{TrDM}^{\mathsf{P}}_h c(X'_{\ell_2-j}) = \mathsf{TrDML}^{\mathsf{P}}_c(X_{\ell-j})$.

Otherwise, if $\ell > \ell_2$ then $X_{\ell-\ell_2+1} = X'_1 = \star \| IV$, and it contradicts that $M$ is free-of-IV-preimage; if $\ell < \ell_2$ then $X'_{\ell_2-\ell+1} = X_1 = \star \| IV$, and $\mathscr{B}_{\rm pre}$ succeeds by outputting $X'_{\ell_2-\ell}$. □

### L.2 Multi-Block Squeezing

When $h > r'$, the involved relations depend on the concrete number of squeezing calls. We take the simplest preimage security as example to illustrate. For simplicity, assume $r' \mid h$, and let $\lambda = h/r'$.

For every $Z = Z[1] \| ... \| Z[\lambda] \in \{0,1\}^h$, we consider a $\lambda$-ary relation $R_{\rm pre2}(Z)$:

$$\begin{aligned}
\big((X_1, ..., X_\lambda), (Y_1, ..., Y_\lambda)\big) \in R_{\rm pre2}(Z) \ \textbf{iff}\ & \mathsf{left}_{r'}(X_1 \oplus Y_1) = Z[1] \\
& \wedge\ \mathsf{left}_{r'}(Y_i) = Z[i] \text{ for all } i \in \{2, ..., \lambda\} \\
& \wedge\ \mathsf{right}_{b-r'}(X_1 \oplus Y_1) = \mathsf{right}_{b-r'}(X_2) \\
& \wedge\ Y_i = X_{i+1} \text{ for all } i \in \{2, ..., \lambda - 1\}. \qquad (82)
\end{aligned}$$

This waters down quantum security bounds that can be obtained by using Theorem 1. As a concrete example, Ascon-DM of Sun et al. [88] has $\lambda = 4$, meaning that the parameter for Theorem 1 has $k = 4$. For $\mathscr{B}^{\mathbf{\Pi},\mathbf{\Pi}^{-1}}$ making $k = 4$ classical queries, it is easy to see $\Pr_{\mathbf{\Pi}}\big[\mathscr{B}^{\mathbf{\Pi},\mathbf{\Pi}^{-1}} \Rightarrow (X_1, ..., X_k) : \big((X_1, ..., X_k), (\mathbf{\Pi}(X_1), ..., \mathbf{\Pi}(X_k))\big) \in R\big]$ cannot be lower than $O(1/2^h)$, since there is a strategy reaching this probability. By these, the generic quantum $R_{\rm pre2}(Z)$-preimage security of Ascon-DM is at best $O(q^8/2^h)$, which is much worse than $O(q^2/2^h)$ of our construction $\mathsf{Sponge\text{-}F}^{\mathsf{P},\mathsf{pd}}$.