# **Polymath: Groth16 Is Not The Limit** Helger Lipmaa, University of Tartu, Estonia

**Crypto 2024 Presentation** 



Computation: fPublic input (statement) XPrivate input (witness) W



#### Computation: fPublic input (statement) X





Computation: *f* Public input (statement) XPrivate input (witness) W



#### Computation: fPublic input (statement) X

STS



Computation: *f* Public input (statement) XPrivate input (witness) W



Computation: fPublic input (statement) XPrivate input (witness) W



Computation: fPublic input (statement) XPrivate input (witness) W



**STS** 

Completeness 

# Computation: *f* Public input (statement) XSTS Knowledge-soundness



Computation: fPublic input (statement) XPrivate input (witness) W



STS

Completeness Knowledge-soundness Zero-knowledge

# Computation: *f* Public input (statement) XSTS



Computation: fPublic input (statement) XPrivate input (witness) W



STS

- Completeness
- Zero-knowledge

Knowledge-soundness • Succinct arguments

STS

#### Computation: f Public input (statement) X



## Landscape







## Landscape



Composition: • GKR + Groth16 • Brakedown + Groth16 • FRI + Groth16



Huge progress in zk-SNARK land in last 5 years Landscape Groth16 still lands supreme after 8 years • Shortest argument Shortest argument Fastest verifier







•  $pp = (p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, [1]_1, [1]_2, \hat{e})$ 

- $pp = (p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, [1]_1, [1]_2, \hat{e})$ 
  - $G_i$  are additive abelian groups of large prime order p

- $pp = (p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, [1]_1, [1]_2, \hat{e})$ 
  - $G_i$  are additive abelian groups of large prime order p
  - $[1]_i$  is a generator of  $\mathbb{G}_i$

- $pp = (p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, [1]_1, [1]_2, \hat{e})$ 
  - $G_i$  are additive abelian groups of large prime order p
  - $[1]_i$  is a generator of  $\mathbb{G}_i$
  - $\hat{e}: \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$  is a bilinear map

- $pp = (p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, [1]_1, [1]_2, \hat{e})$ 
  - $G_i$  are additive abelian groups of large prime order p
  - $[1]_i$  is a generator of  $\mathbb{G}_i$
  - $\hat{e}: \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$  is a bilinear map
  - $\hat{e}([a]_1, [b]_2) = [ab]_T \text{ for } a, b \in \mathbb{F}_p$



- $pp = (p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, [1]_1, [1]_2, \hat{e})$ 
  - $G_i$  are additive abelian groups of large prime order p
  - $[1]_i$  is a generator of  $\mathbb{G}_i$
  - $\hat{e}: \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$  is a bilinear map
  - $\hat{e}([a]_1, [b]_2) = [ab]_T$  for  $a, b \in \mathbb{F}_p$
- "Standard" curve for 128-bit security level: BLS12-381

- $pp = (p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, [1]_1, [1]_2, \hat{e})$ 
  - $G_i$  are additive abelian groups of large prime order p
  - $[1]_i$  is a generator of  $\mathbb{G}_i$
  - $\hat{e}: \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$  is a bilinear map
  - $\hat{e}([a]_1, [b]_2) = [ab]_T \text{ for } a, b \in \mathbb{F}_p$
- "Standard" curve for 128-bit security level: BLS12-381
  - $\ell(\mathbb{F}_p) = 256$ ,  $\ell(\mathbb{G}_1) = 384$ ,  $\ell(\mathbb{G}_2) = 768$  (bits)

- $pp = (p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, [1]_1, [1]_2, \hat{e})$ 
  - $G_i$  are additive abelian groups of large prime order p
  - $[1]_i$  is a generator of  $\mathbb{G}_i$
  - $\hat{e}: \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$  is a bilinear map
  - $\hat{e}([a]_1, [b]_2) = [ab]_T \text{ for } a, b \in \mathbb{F}_p$
- "Standard" curve for 128-bit security level: BLS12-381
  - $\ell(\mathbb{F}_p) = 256, \, \ell(\mathbb{G}_1) = 384, \, \ell(\mathbb{G}_2) = 768 \, \text{(bits)}$
- Curves for 192-bit security level:

- $pp = (p, \mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, [1]_1, [1]_2, \hat{e})$ 
  - $G_i$  are additive abelian groups of large prime order p
  - $[1]_i$  is a generator of  $\mathbb{G}_i$
  - $\hat{e}: \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$  is a bilinear map
  - $\hat{e}([a]_1, [b]_2) = [ab]_T \text{ for } a, b \in \mathbb{F}_p$
- "Standard" curve for 128-bit security level: BLS12-381

•  $\ell(\mathbb{F}_p) = 256, \, \ell(\mathbb{G}_1) = 384, \, \ell(\mathbb{G}_2) = 768 \, \text{(bits)}$ 

- Curves for 192-bit security level:
  - $\ell(\mathbb{F}_p) = 256, \ \ell(\mathbb{G}_1) = 512, \ \ell(\mathbb{G}_2) = 2048$  (bits)

# Groth16: Bird's-Eye

 $\operatorname{srs}(f)$ 

Computation: fPublic input (statement) XPrivate input (witness) W



SRS depends on the circuit



View

 $\operatorname{srs}(f)$ 

# Groth16: Bird's-Eye

Computation: fPublic input (statement) XPrivate input (witness) W



SRS depends on the circuit • Argument length: only 3 group elements



# Groth16: Bird's-Eye

Computation: fPublic input (statement) XPrivate input (witness) W



- SRS depends on the circuit
- Argument length: only 3 group elements
- Verifier executes three pairings and X group ops



• Groth16 has three group elements



- Groth16 has three group elements
- Lower bound [Groth, EC16]:



- Groth16 has three group elements
- Lower bound [Groth, EC16]: lacksquare
  - At least two group elements needed



- Groth16 has three group elements
- Lower bound [Groth, EC16]:
  - At least two group elements needed
  - One of them has to be in  $\mathbb{G}_2$



- Groth16 has three group elements
- Lower bound [Groth, EC16]:
  - At least two group elements needed
  - One of them has to be in  $\mathbb{G}_2$
- Using a different arithmetization, one can have two group elements



- Groth16 has three group elements
- Lower bound [Groth, EC16]:
  - At least two group elements needed
  - One of them has to be in  $\mathbb{G}_{2}$
- Using a different arithmetization, one can have two group elements
  - SAP (Square Arithmetic Programming) instead of R1CS



- Groth16 has three group elements
- Lower bound [Groth, EC16]:
  - At least two group elements needed
  - One of them has to be in  $\mathbb{G}_{2}$
- Using a different arithmetization, one can have two group elements
  - SAP (Square Arithmetic Programming) instead of R1CS
  - However, then one relies on symmetric pairings



- Groth16 has three group elements
- Lower bound [Groth, EC16]:
  - At least two group elements needed
  - One of them has to be in  $\mathbb{G}_{2}$
- Using a different arithmetization, one can have two group elements SAP (Square Arithmetic Programming) instead of R1CS • However, then one relies on symmetric pairings Group elements will be considerably longer => worse in practice



- Groth16 has three group elements
- Lower bound [Groth, EC16]:
  - At least two group elements needed
  - One of them has to be in  $\mathbb{G}_2$
- Using a different arithmetization, one can have two group elements
  - SAP (Square Arithmetic Programming) instead of R1CS
  - However, then one relies on symmetric pairings
  - Group elements will be considerably longer => worse in practice

The lower bound works in the standard model (no RO)



- Groth16 has three group elements
- Lower bound [Groth, EC16]:
  - At least two group elements needed
  - One of them has to be in  $\mathbb{G}_{2}$
- Using a different arithmetization, one can have two group elements
  - SAP (Square Arithmetic Programming) instead of R1CS
  - However, then one relies on symmetric pairings
  - Group elements will be considerably longer => worse in practice

It talks about **#group elements**, not **bit-length** 

# The lower bound works in the standard model (no RO)





# Scenic Route to Polymath For non-muggles

• Problem:  $\mathbb{G}_2$  elements are long




- Problem:  $\mathbb{G}_2$  elements are long
- $[b]_2 \Longrightarrow [b]_1$ , but how?





- Problem:  $\mathbb{G}_2$  elements are long
- $[b]_2 \Longrightarrow [b]_1$ , but how?
- Groth16 uses pairings to do quadratic checks



- Problem:  $\mathbb{G}_2$  elements are long
- $[b]_2 \Longrightarrow [b]_1$ , but how?
- Groth16 uses pairings to do quadratic checks
- We can KZG-open the polynomial commitment  $[b]_1$  to some  $\overline{b}$  and do quadratic checks by using  $\overline{b}$



# tic checks commitment $[b]_1$ to some $\overline{b}$ and do

- Problem:  $\mathbb{G}_2$  elements are long
- $[b]_2 \Longrightarrow [b]_1$ , but how?
- Groth16 uses pairings to do quadratic checks
- We can KZG-open the polynomial commitment  $[b]_1$  to some b and do quadratic checks by using b
- KZG opening is shorter than a  $\mathbb{G}_{2}$  element



- Problem:  $\mathbb{G}_2$  elements are long
- $[b]_2 \Longrightarrow [b]_1$ , but how?
- Groth16 uses pairings to do quadratic checks We can KZG-open the polynomial commitment  $[b]_1$  to some  $\bar{b}$  and do quadratic checks by using  $\bar{b}$ 
  - KZG opening is shorter than a  $\mathbb{G}_2$  element
  - (a field element  $\overline{b}$  and a  $\mathbb{G}_1$  element  $[h]_1$ )



- Problem:  $\mathbb{G}_2$  elements are long
- $[b]_2 \Longrightarrow [b]_1$ , but how?
- Groth16 uses pairings to do quadratic checks We can KZG-open the polynomial commitment  $[b]_1$  to some b and do quadratic checks by using b
  - KZG opening is shorter than a  $\mathbb{G}_2$  element
  - (a field element  $\overline{b}$  and a  $\mathbb{G}_1$  element  $[h]_1$ )

Problem:

- we still have  $[b]_1$  in the argument!



•  $\ell([b]_2) < \ell([b]_1) + \ell(\bar{b}) + \ell([h]_1)$  in 128-bit level

If we use SAP instead of R1CS, we get  $[b]_1 = [a]_1$ 





If we use SAP instead of R1CS, we get  $[b]_1 = [a]_1$ 

• No need to send it!



If we use SAP instead of R1CS, we get  $[b]_1 = [a]_1$ 

- No need to send it!
- **Cost:** circuit  $\approx 2$  longer => slower prover



If we use SAP instead of R1CS, we get  $[b]_1 = [a]_1$ 

- No need to send it!
- **Cost:** circuit  $\approx 2$  longer => slower prover  $\bullet$ 
  - Multiplication gates => squaring gates



If we use SAP instead of R1CS, we get  $[b]_1 = [a]_1$ 

- No need to send it!
- **Cost:** circuit  $\approx 2$  longer => slower prover
  - Multiplication gates => squaring gates

Problem:

- Not clear how to use KZG



• Groth16 has five trapdoors, KZG is univariate

• Univariatization:





- Univariatization:
  - Replace each trapdoor with  $\chi^i$  for some *i* and a single trapdoor  $\chi$



- Univariatization:
  - Replace each trapdoor with  $x^i$  for some *i* and a single trapdoor x
- Also done in [Lipmaa, PKC 2022] who used exhaustive search to find i's



- Univariatization:
  - Replace each trapdoor with  $\chi^{i}$  for some *i* and a single trapdoor  $\chi$
- Also done in [Lipmaa, PKC 2022] who used exhaustive search to find i's

### Problem:

- even after exhaustive search, the exponents i are quite large KZG prover time  $\Omega$ (polynomial degree) => Results in high prover complexity

![](_page_50_Picture_8.jpeg)

**Observation 1:** Groth16 for SAP has -1 trapdoor 

![](_page_51_Picture_2.jpeg)

![](_page_51_Picture_3.jpeg)

Observation 1: Groth16 for SAP has -1 trapdoor
Observation 2:

![](_page_52_Picture_2.jpeg)

![](_page_52_Picture_3.jpeg)

- **Observation 1:** Groth16 for SAP has -1 trapdoor **Observation 2:** 
  - One trapdoor in Groth16 is only needed to mask SRS elements corresponding to the statement

![](_page_53_Picture_3.jpeg)

- **Observation 1:** Groth16 for SAP has -1 trapdoor **Observation 2:** 
  - One trapdoor in Groth16 is only needed to mask SRS elements corresponding to the statement

![](_page_54_Picture_4.jpeg)

- **Observation 1:** Groth16 for SAP has -1 trapdoor **Observation 2:** 
  - One trapdoor in Groth16 is only needed to mask SRS elements corresponding to the statement
  - - Verifier becomes faster

![](_page_55_Picture_5.jpeg)

- **Observation 1:** Groth16 for SAP has -1 trapdoor **Observation 2:** 
  - One trapdoor in Groth16 is only needed to mask SRS elements corresponding to the statement

Verifier becomes faster

• In Polymath, V interpolates a polynomial in  $\mathbb{F}_p$  of degree  $\mathbb{X}$ 

![](_page_56_Picture_6.jpeg)

- **Observation 1:** Groth16 for SAP has -1 trapdoor **Observation 2:** 
  - One trapdoor in Groth16 is only needed to mask SRS elements corresponding to the statement

Verifier becomes faster

- In Polymath, V interpolates a polynomial in  $\mathbb{F}_p$  of degree  $\mathbb{X}$ 
  - Instead of doing |X| -long MSM in Groth16

![](_page_57_Picture_7.jpeg)

• We only have three trapdoors

![](_page_58_Picture_2.jpeg)

![](_page_58_Picture_3.jpeg)

- We only have three trapdoors
- It is easier to choose "good" exponents!

![](_page_59_Picture_3.jpeg)

![](_page_59_Picture_4.jpeg)

- We only have three trapdoors
- It is easier to choose "good" exponents! We replace two trapdoors with  $y^{\alpha}$  and  $y^{\gamma}$ , where  $y = x^{\sigma}$

![](_page_60_Picture_3.jpeg)

- We only have three trapdoors
- It is easier to choose "good" exponents! We replace two trapdoors with  $y^{\alpha}$  and  $y^{\gamma}$ , where  $y = x^{\sigma}$ 
  - x is real trapdoor,  $\alpha, \gamma, \sigma$  are field elements

![](_page_61_Picture_4.jpeg)

- We only have three trapdoors
- It is easier to choose "good" exponents! We replace two trapdoors with  $y^{\alpha}$  and  $y^{\gamma}$ , where  $y = x^{\sigma}$ 
  - x is real trapdoor,  $\alpha, \gamma, \sigma$  are field elements
- Exhaustive search to find small exponents that result in knowledgesoundness

![](_page_62_Picture_5.jpeg)

- We only have three trapdoors
- It is easier to choose "good" exponents! We replace two trapdoors with  $y^{\alpha}$  and  $y^{\gamma}$ , where  $y = x^{\sigma}$ 
  - x is real trapdoor,  $\alpha, \gamma, \sigma$  are field elements
- Exhaustive search to find small exponents that result in knowledgesoundness
- Kamek saith:  $\alpha = 3$ ,  $\gamma = -5$ ,  $\sigma = n + 3$

![](_page_63_Picture_6.jpeg)

![](_page_63_Picture_7.jpeg)

- We only have three trapdoors
- It is easier to choose "good" exponents! We replace two trapdoors with  $y^{\alpha}$  and  $y^{\gamma}$ , where  $y = x^{\sigma}$ 
  - x is real trapdoor,  $\alpha, \gamma, \sigma$  are field elements
- Exhaustive search to find small exponents that result in knowledgesoundness
- Kamek saith:  $\alpha = 3$ ,  $\gamma = -5$ ,  $\sigma = n + 3$

Problem:

- SRS is circuit-dependent
- It does not contain enough elements to compute  $[h]_1$

![](_page_64_Picture_9.jpeg)

![](_page_64_Picture_10.jpeg)

This adds elements to the SRS

![](_page_65_Picture_3.jpeg)

### We add another trapdoor $[z]_1$ that is **only** used to compute KZG opening

This adds elements to the SRS

![](_page_66_Picture_3.jpeg)

![](_page_66_Picture_4.jpeg)

### We add another trapdoor $[z]_1$ that is **only** used to compute KZG opening

Completeness and zero-knowledge proofs are short and standard

- Completeness and zero-knowledge proofs are short and standard
- Soundness proof is seven pages

- Completeness and zero-knowledge proofs are short and standard
- Soundness proof is seven pages
- 1. Groth16 itself has a complicated proof

- Completeness and zero-knowledge proofs are short and standard
- Soundness proof is seven pages
- 1. Groth16 itself has a complicated proof
- 2. Using virtual trapdoors  $x^i$  instead of independent trapdoors makes proof more complicated

- Completeness and zero-knowledge proofs are short and standard
- Soundness proof is seven pages
- 1. Groth16 itself has a complicated proof
- 2. Using virtual trapdoors  $x^i$  instead of independent trapdoors makes proof more complicated
- 3. To get tight security after Fiat-Shamir, we prove special-soundness
- Completeness and zero-knowledge proofs are short and standard
- Soundness proof is seven pages
- 1. Groth16 itself has a complicated proof
- 2. Using virtual trapdoors  $\chi^{l}$  instead of independent trapdoors makes proof more complicated
- 3. To get tight security after Fiat-Shamir, we prove special-soundness 4. Proof is in AGM with Oblivious Sampling [Lipmaa-Parisella-Siim, TCC 2023]



- Completeness and zero-knowledge proofs are short and standard
- Soundness proof is seven pages
- 1. Groth16 itself has a complicated proof
- 2. Using virtual trapdoors  $\chi^{l}$  instead of independent trapdoors makes proof more complicated
- LPS23 noted that KZG is often used wrongly
- 3. To get tight security after Fiat-Shamir, we prove special-soundness 4. Proof is in AGM with Oblivious Sampling [Lipmaa-Parisella-Siim, TCC 2023]



- Completeness and zero-knowledge proofs are short and standard
- Soundness proof is seven pages
- 1. Groth16 itself has a complicated proof
- 2. Using virtual trapdoors  $\chi^{l}$  instead of independent trapdoors makes proof more complicated
- 3. To get tight security after Fiat-Shamir, we prove special-soundness 4. Proof is in AGM with Oblivious Sampling [Lipmaa-Parisella-Siim, TCC 2023] LPS23 noted that KZG is often used wrongly
- - Constructions are secure in AGM but not when adversary can do o.s.



- Completeness and zero-knowledge proofs are short and standard
- Soundness proof is seven pages
- 1. Groth16 itself has a complicated proof
- 2. Using virtual trapdoors  $\chi^{l}$  instead of independent trapdoors makes proof more complicated
- 3. To get tight security after Fiat-Shamir, we prove special-soundness 4. Proof is in AGM with Oblivious Sampling [Lipmaa-Parisella-Siim, TCC 2023] LPS23 noted that KZG is often used wrongly
- - Constructions are secure in AGM but not when adversary can do o.s.
  - And o.s. is for free!



- Completeness and zero-knowledge proofs are short and standard
- Soundness proof is seven pages
- 1. Groth16 itself has a complicated proof
- 2. Using virtual trapdoors  $\chi^{l}$  instead of independent trapdoors makes proof more complicated
- 3. To get tight security after Fiat-Shamir, we prove special-soundness 4. Proof is in AGM with Oblivious Sampling [Lipmaa-Parisella-Siim, TCC 2023] LPS23 noted that KZG is often used wrongly
- - Constructions are secure in AGM but not when adversary can do o.s.
  - And o.s. is for free!



**Part** of Polymath's proof is machine-checked



• First improvement on Groth16 since 2016





- First improvement on Groth16 since 2016



# Hopefully encourages other researchers to improve Groth16 even further

- First improvement on Groth16 since 2016
- Pros:



# Hopefully encourages other researchers to improve Groth16 even further

- First improvement on Groth16 since 2016
- Pros:
  - Smaller arguments (more prominent in 192/256-bit security level)





- First improvement on Groth16 since 2016
- Pros:

  - Smaller arguments (more prominent in 192/256-bit security level) • No adversarially created  $\mathbb{G}_2$  elements — good for batching





- First improvement on Groth16 since 2016
- Pros:

  - Smaller arguments (more prominent in 192/256-bit security level) • No adversarially created  $\mathbb{G}_2$  elements — good for batching • Verifier is potentially faster for a short public input





- First improvement on Groth16 since 2016
- Pros:

  - Smaller arguments (more prominent in 192/256-bit security level) • No adversarially created  $\mathbb{G}_2$  elements — good for batching • Verifier is potentially faster for a short public input
  - - And definitely faster for a long public input





- First improvement on Groth16 since 2016 Hopefully encourages other researchers to improve Groth16 even further
- Pros:

  - Smaller arguments (more prominent in 192/256-bit security level) • No adversarially created  $\mathbb{G}_2$  elements — good for batching
  - Verifier is potentially faster for a short public input
    - And definitely faster for a long public input
- Cons:

#### Mitigated in SNARK composition when used as a final SNARK





- First improvement on Groth16 since 2016
  - Hopefully encourages other researchers to improve Groth16 even further
- Pros:

  - Smaller arguments (more prominent in 192/256-bit security level) • No adversarially created  $\mathbb{G}_2$  elements — good for batching • Verifier is potentially faster for a short public input
- And definitely faster for a long public input
- Cons:
  - Prover is slower



 Mitigated in SNARK composition when used as a final SNARK • Prover's input is shorter => prover speed less important



- First improvement on Groth16 since 2016
- Pros:
  - Smaller arguments (more prominent in 192/256-bit security level)
  - No adversarially created  $\mathbb{G}_2$  elements good for batching
  - Verifier is potentially faster for a short public input
    - And definitely faster for a long public input
- **Cons:** 
  - Prover is slower
  - Uses random oracle model on top of AGM(OS)

 Mitigated in SNARK composition when used as a final SNARK • Prover's input is shorter => prover speed less important Any known reasonable initial SNARK candidate uses ROM



# Hopefully encourages other researchers to improve Groth16 even further