hexagon:2611.00157v1

AN EXPLICIT PRIME GAP UNDER THE RIEMANN HYPOTHESIS, FORMALIZED IN LEAN WITH ITS PRIME POWER INPUT MADE ELEMENTARY

STEVE HUBBARD

Abstract. We report on a Lean 4 formalization of Theorem 2 of Dudek, 2015, as numbered in arXiv v2, which states that if the Riemann hypothesis is true, then for every real $x \geq 2$ there is a prime in the interval $(x - (4/\pi)\sqrt{x}\log x, x]$. In the formalization, Theorem 2 is proved assuming the Riemann hypothesis and one cited explicit zero-counting bound of Hasanalizade, Shen and Wong, arXiv:2107.06506. Four cited inputs of the printed proof are replaced in the formal proof. The two that carry the argument are these: first, the prime power contribution $\psi - \theta$ is required only across the moving window $(x - h, x + h]$ and is proved outright in Lean by an elementary counting argument with no cited input and without assuming the Riemann hypothesis; second, the sum over nontrivial zeros is bounded from the cited Hasanalizade, Shen and Wong zero count for $T \geq 70$ by trading slack between the head and tail sums at the split point $T = 2x/h$. The Riemann hypothesis is open and this note makes no progress on it.

1. The paper

This note documents a formalization in Lean 4 of the main explicit prime gap theorem from A. W. Dudek, *On the Riemann hypothesis and the difference between primes*, Int. J. Number Theory 11, 2015, no. 3, 771–778, doi:10.1142/S1793042115500426; preprint arXiv:1402.6417, v2, 20 May 2014 [Dudek]. That paper improves the constant 8/5 of Ramaré and Saouter [RaSa, Theorem 1] to $4/\pi$. Later work went further: Dudek, Grenié and Molteni [DGM, Theorem 1.1] proved a prime in a symmetric interval of total length $(1+4/\log x)\sqrt{x}\log x$, and Carneiro, Milinovich and Soundararajan [CMS, Theorem 5] proved, under the Riemann hypothesis, that for $x \geq 4$ there is a prime in $[x, x + (22/25)\sqrt{x}\log x]$. Theorem 2 of [Dudek] is therefore not the best known constant, and this note is about the formal record of its proof. It is written for readers who use or check formal proofs in analytic number theory.

All theorem, lemma and page numbers of [Dudek] in this note are those of arXiv v2. The journal numbering may differ: [CMS] cites the journal paper with Theorem 1.3 for Theorem 3 of arXiv v2.

The primary focus of this note is Theorem 2 of arXiv v2 of [Dudek]:

Theorem 1.1 (Dudek [Dudek], Theorem 2 of arXiv v2)**.** *Suppose the Riemann hypothesis is true. Then there is a prime in the interval*

$$\left(x - \frac{4}{\pi}\sqrt{x}\log x,\ x\right]$$

for all $x \geq 2$.

In addition to Theorem 2, Theorem 1 of [Dudek], which Dudek quotes from Cramér, gives a constant $c > 0$ such that

$$\pi(x + c\sqrt{x}\log x) - \pi(x) > \sqrt{x}$$

for all $x \geq 2$, and Dudek's Theorem 3 gives a prime in the interval

$$\big(x - (1+\varepsilon)\sqrt{x}\log x,\, x\big]$$

for all sufficiently large x. The same Lean entry also proves Theorems 1 and 3 under the same two hypotheses, with $c = 10^8$ for Theorem 1, whereas the remark of the paper that $c = 3 + \varepsilon$ works for large x is not part of the entry.

The purpose of this note is to record which inputs the formal proof took and how they differ from those printed in [Dudek]. No discrepancy was found in the steps of Dudek's argument that were formalized. Theorem 2 is proved assuming the Riemann hypothesis and one cited zero count. The Riemann hypothesis is open and this note makes no progress on it. The binary Goldbach conjecture also remains open and is not treated here.

The mathematical content of this note is two observations. The proof uses $\psi - \theta$ only across a short window, where the prime power mass is $O(\log^2 x)$ by counting; this is a simplification of this proof over citing Rosser and Schoenfeld. The zero sum can be closed with a zero count whose constant is too large for the separate head and tail bounds, by trading slack at the split point; this is a convenience forced by the chosen input. Neither is claimed as new. The rest of the Lean entry is formalization work.

Outline of the paper's proof. Section 2 of [Dudek] establishes Theorem 2 using a weight w on the open interval $(x-h, x+h)$ together with the integrated Chebyshev function

$$\psi_1(x) = \int_2^x \psi(t)\,dt.$$

The argument in Section 2 of [Dudek] proceeds through the following steps:

(1) **Explicit formula for ψ_1, Lemma 4 of** [Dudek]**.** For $x > 0$, with x not an integer,

$$\psi_1(x) = \frac{x^2}{2} - \sum_\rho \frac{x^{\rho+1}}{\rho(\rho+1)} - x\log(2\pi) + \varepsilon(x), \qquad |\varepsilon(x)| < \frac{12}{5},$$

where the sum runs over the nontrivial zeros ρ of the Riemann zeta function. The printed argument bounds a sum over zeros by $2^{3/2}\sum_\rho 1/|\rho|^2$ and quotes the value of this sum from Davenport [Da].

(2) **Weighted prime-power sum via second differences, Lemma 5 of** [Dudek]**.** For $x > 0$ and $h > 0$,

$$\sum_n \Lambda(n)w(n) = h - \frac{1}{h}\Sigma + \varepsilon(h), \qquad |\varepsilon(h)| < \frac{48}{5h},$$

where Λ is the von Mangoldt function and

$$\Sigma = \sum_\rho \frac{(x+h)^{\rho+1} - 2x^{\rho+1} + (x-h)^{\rho+1}}{\rho(\rho+1)}.$$

(3) **Splitting the sum over zeros.** Writing $\rho = 1/2 + i\gamma$ under the Riemann hypothesis, the sum Σ over nontrivial zeros is split at $|\gamma| = \alpha x/h$. To bound the head of the sum, the paper uses a zero count, specifically

$$N(T) < \frac{T \log T}{2\pi} \qquad \text{for } T > 15,$$

cited as Corollary 1 of Trudgian, Math. Comp. 81, 2012 [Tr]. To bound the tail of the sum, the paper uses Lemma 1 (ii) of Skewes [Sk].

(4) **Passing from Λ to primes.** To pass from the von Mangoldt function Λ to primes, the paper uses, in the paper's words, Theorem 14 and equation 5.5 of its reference [7], Rosser and Schoenfeld, 1962 [RS], namely the two-sided bound

$$0.98\sqrt{x} < \psi(x) - \theta(x) < 1.11\sqrt{x} + 3x^{1/3} \qquad \text{for } x \geq 121.$$

(5) **Positivity of the lower bound for $x \geq 65000$ and finite verification.** Choosing $\alpha = 2$ and

$$h = \frac{2}{\pi}\sqrt{x}\log x,$$

the display at the foot of page 6 of arXiv v2 of [Dudek] states that $\sum_{x-h<p\leq x+h} \log p$, the sum of $\log p$ over primes p in $(x-h, x+h]$, exceeds

$$\begin{aligned}\mathrm{rhs}(x) = h - \frac{2}{\pi}\left(\frac{x}{\sqrt{x-h}} + \frac{(x+h)^{3/2}}{x}\right)\log\left(\frac{2x}{h}\right) \\ - 1.11\sqrt{x+h} - 3(x+h)^{1/3} + 0.98\sqrt{x-h} - \frac{48}{5h},\end{aligned}$$

which is positive for all $x \geq 65000$. Positivity of $\mathrm{rhs}(x)$ guarantees the existence of a prime in $(x-h, x+h]$. Replacing $x+h$ by x then yields Theorem 2 from about 66798.7, namely

$$65000 + \frac{2}{\pi}\sqrt{65000}\log 65000.$$

The remaining values of x below this threshold were checked by computer in the paper, using Mathematica, with no table printed.

2. What the formal proof does

In the Lean 4 formalization [Palomar, Repo], the main theorem corresponding to Theorem 2 of [Dudek] has the full name

`LeanLab.PNTPort.ExplicitPsi1.Trivial.theorem2_of_RH_hsw`.

Its exact Lean declaration, together with the interval predicate and width function, is as follows:

```
theorem theorem2_of_RH_hsw (hRH : RiemannHypothesis)
    (hN : LeanLab.Dudek2014.SigmaH.HSWN70 gZ) :
    ∀ x : ℝ, 2 ≤ x → LeanLab.Dudek2014.PrimeInDudekInterval x
```

with

```
def PrimeInDudekInterval (x : ℝ) : Prop :=
  ∃ p : ℕ, p.Prime ∧ x - dudekH x < p ∧ (p : ℝ) ≤ x
noncomputable def dudekH (x : ℝ) : ℝ := 4 / π * √x * Real.log x
```

The hypotheses and objects appearing in this statement are defined as follows:

- `RiemannHypothesis` is the standard statement of the Riemann hypothesis from Mathlib [Mathlib].
- `gZ` is the family of imaginary parts of the zeros of `riemannZeta` with real part at least $-1/2$, which are the nontrivial zeros, indexed with multiplicity given by `analyticOrderNatAt`.
- `HSWN70 gZ` is the one cited input, included explicitly as a hypothesis to the theorem and never declared as an axiom.

The predicate `HSWN70` is defined in Lean by:

```
def HSWN70 (γ : ι → ℝ) : Prop :=
  (∀ T : ℝ, 70 ≤ T → ∀ s : Finset ι,
    ((s.filter (fun i => |γ i| ≤ T)).card : ℝ) ≤ 2 * (Mt T + Rt T)) ∧
  (∀ T t : ℝ, 70 ≤ T → T ≤ t → ∀ s : Finset ι,
    ((s.filter (fun i => T < |γ i| ∧ |γ i| ≤ t)).card : ℝ) ≤
      2 * ((Mt t + Rt t) - (Mt T - Rt T)))
```

where the main term $\mathrm{Mt}(T)$ and remainder bound $\mathrm{Rt}(T)$ are given by

$$\mathrm{Mt}(T) = \frac{T}{2\pi}\log\left(\frac{T}{2\pi e}\right) + \frac{7}{8},$$
$$\mathrm{Rt}(T) = 0.1038\log T + 0.2573\log\log T + 10.2425.$$

This hypothesis encodes Corollary 1.2 of Hasanalizade, Shen and Wong, arXiv:2107.06506 [HSW], which states that

$$\left|N(T) - \frac{T}{2\pi}\log\left(\frac{T}{2\pi e}\right)\right| \le 0.1038\log T + 0.2573\log\log T + 9.3675 \qquad \text{for } T \ge e.$$

In the Lean formalization, this bound is used only for $T \ge 70$. The Lean definitions add $7/8$ to the main term and $7/8$ to the remainder, so that

$$9.3675 + \frac{7}{8} = 10.2425.$$

Thus $\mathrm{Mt}(T) - \mathrm{Rt}(T)$ is exactly the printed lower bound for $N(T)$, and $\mathrm{Mt}(T) + \mathrm{Rt}(T)$ is the printed upper bound plus $7/4$. The factor 2 counts ordinates γ of both signs, $|\gamma| \le T$ and $T < |\gamma| \le t$, since $N(T)$ counts only $0 < \gamma \le T$. Zeros are counted with multiplicity, the standard reading of $N(T)$, which [HSW] prints as a set count. Consequently, the hypothesis `HSWN70 gZ` holds for the actual nontrivial zeros of the Riemann zeta function given Corollary 1.2 of [HSW], ensuring that `theorem2_of_RH_hsw` is not vacuous. The cited zero count itself is not proved in Lean.

Table 1 summarizes the relationship between the printed proof in [Dudek] and the formal proof. Four cited inputs of the printed proof are replaced: the zero count of [Tr], the bound of [RS], the tail bound of [Sk] and the value from [Da]. The zero count of [HSW] is a choice of input, not a correction; this note does not examine whether the input of [Tr] also suffices. The smaller differences are these. The Lean definition of ψ_1 is the sum $\sum_{n \le y}(y-n)\Lambda(n)$, which has the same value as the printed integral for $y \ge 2$. The Lean form of Lemma 5 adds the side condition $h < x$, which holds for $x \ge 65000$. Lemma 4 is proved by its own route, and integer y is allowed. Zeros are counted with multiplicity. We present the mathematics of the prime power step and the zero sum in the following two subsections.

TABLE 1. Comparison between the inputs and verification steps printed in Dudek [Dudek] and those used in the Lean 4 formalization.

Component	Printed paper [Dudek]	Lean 4 formalization
Zero hypothesis	Riemann hypothesis	`RiemannHypothesis` (Mathlib)
Zero count input	Trudgian [Tr], $T > 15$	`HSWN70 gZ` from [HSW], $T \geq 70$
Prime power step	Cited from [RS] for $x \geq 121$	`window_holds`, proved outright
ψ_1 error bound	$\lvert\varepsilon(x)\rvert < 12/5$	`error_bound` $\leq 2.393 < 12/5$
Sum $\sum_\rho 1/\lvert\rho\rvert^2$	Value from Davenport [Da]	Not used, the route of `error_bound` differs
Tail of $\sum 1/\gamma^2$	Lemma 1 (ii) of Skewes [Sk]	`tail_inv_sq`, from `HSWN70`
Large range	$x \geq 65000$ ($X \gtrsim 66798.7$)	`rhs_pos` ($x \geq 65000$), $X \geq 66799$
Small range	Mathematica check (unprinted)	`small_range` ($2 \leq x < 67001$, kernel)

2.1. **The prime power step, proved with no cited input.** A key observation of the formalization is that the proof of Theorem 2 uses the difference $\psi - \theta$ only through its increment across the window $(x-h, x+h]$. In Lean, this windowed inequality is named `SchoenfeldWindow` and is proved as the theorem `LeanLab.PrimePowerWindow.window_holds`:

```
def SchoenfeldWindow : Prop :=
  ∀ x : ℝ, 65000 ≤ x →
    (psiF (x + hD x) - thetaF (x + hD x)) -
      (psiF (x - hD x) - thetaF (x - hD x)) <
      1.11 * √(x + hD x) + 3 * (x + hD x) ^ (1 / 3 : ℝ) -
        0.98 * √(x - hD x)
```

where

$$\mathrm{hD}(x) = \frac{2}{\pi}\sqrt{x}\log x.$$

The theorem `window_holds` is proved outright in Lean, assuming no Riemann hypothesis and using no cited input. A companion lemma, `schoenfeld_window_of`, records that the two-sided bound of Rosser and Schoenfeld [RS] cited in the paper implies `SchoenfeldWindow`; thus the window form is a logical weakening of the printed input and is all that Dudek's argument requires.

Proposition 2.1 (`LeanLab.PrimePowerWindow.window_holds`)**.** *For every real $x \geq 65000$, with $h = \mathrm{hD}(x) = (2/\pi)\sqrt{x}\log x$,*

$$\big(\psi(x+h)-\theta(x+h)\big)-\big(\psi(x-h)-\theta(x-h)\big) < 1.11\sqrt{x+h}+3(x+h)^{1/3}-0.98\sqrt{x-h}.$$

The proof goes through the following bound on the mass of the window.

Lemma 2.2 (`LeanLab.PrimePowerWindow.mass_le` with `diff_eq`)**.** *Let $1 \leq a \leq b$ be real numbers with $b^{1/3} < a^{1/3} + 1$. Then*

$$\big(\psi(b) - \theta(b)\big) - \big(\psi(a) - \theta(a)\big) \leq \frac{\log b}{2}\big(\sqrt{b} - \sqrt{a} + 1\big) + \frac{(\log b)^2}{3\log 2}.$$

Proof. Since $\psi(y) = \sum_{n \leq y} \Lambda(n)$ and $\theta(y) = \sum_{p \leq y} \log p$, the difference $\psi(y) - \theta(y)$ is the sum of $\Lambda(n)$ over non-prime integers $n \leq y$. Consequently, the left side of

the lemma is precisely the von Mangoldt mass of the prime powers p^k with $k \geq 2$ lying in the interval $(a, b]$.

First, the lemma `sum_nonprime_le` bounds the non-prime von Mangoldt mass over a natural-number interval $(A, B]$ with $A \leq B$:

$$\sum_{\substack{A<n\leq B \\ n \text{ not prime}}} \Lambda(n) \leq \sum_{2\leq k\leq \log_2 B} \frac{\log B}{k} \cdot \#\{m \in \mathbb{N} : A < m^k \leq B\}.$$

Indeed, every non-prime $n \in (A, B]$ with $\Lambda(n) \neq 0$ is of the form $n = p^k$ for a prime p and an integer $k \geq 2$ satisfying $k \leq \log_2 B$. For such n, its von Mangoldt weight satisfies

$$\Lambda(n) = \log p = \frac{\log(p^k)}{k} \leq \frac{\log B}{k},$$

and for each fixed exponent k, the map $n \mapsto p$ is injective.

Second, the lemma `count_le` bounds the number of k-th powers in the integer interval $(\lfloor a \rfloor, \lfloor b \rfloor]$ for real numbers $0 \leq a \leq b$:

$$\#\{m \in \mathbb{N} : \lfloor a \rfloor < m^k \leq \lfloor b \rfloor\} \leq b^{1/k} - a^{1/k} + 1.$$

Applying these two bounds separates the prime powers in $(a, b]$ into squares, $k = 2$, and higher powers, $3 \leq k \leq \log_2 b$:

- For $k = 2$, there are at most $\sqrt{b} - \sqrt{a} + 1$ squares in $(a, b]$, each carrying weight $\log p \leq \frac{1}{2} \log b$.
- For each $3 \leq k \leq \log_2 b$, the bound $b^{1/k} < a^{1/k} + 1$ follows from the case $k = 3$ by `rpow_sub_mono`, so there is at most one k-th power in $(a, b]$, by `count_le_one`, carrying weight $\log p \leq \frac{1}{k} \log b \leq \frac{1}{3} \log b$.

There are at most $\log b / \log 2$ exponents k, which gives the last term. □

Proof of Proposition 2.1. Let $h = \mathrm{hD}(x)$, $a = x - h$, $b = x + h$, $u = x^{1/6}$, $L = \log x$ and $M = \log b$. For all $x \geq 65000$,

$$u \geq 6.34, \qquad L \leq 1.5u + 2.32, \qquad L \geq 11, \qquad h \leq 0.63662\, u^3 L.$$

These give $h \leq x/10$, so $0.9x \leq a$ and $b \leq 1.1x$. Hence $\sqrt{a} \geq 0.948\, u^3$, and $(\sqrt{b} - \sqrt{a})(\sqrt{b} + \sqrt{a}) = 2h$ gives $\sqrt{b} - \sqrt{a} \leq 0.6716\, L$. Also $a^{1/3} \geq 0.965\, u^2$ gives $b < (a^{1/3} + 1)^3$, so Lemma 2.2 applies. Further $L \leq M \leq L + 0.1 \leq 1.5u + 2.42$. Since $\sqrt{a} \leq u^3 \leq \sqrt{b}$ and $u^2 \leq b^{1/3}$, the target satisfies

$$1.11\sqrt{b} + 3b^{1/3} - 0.98\sqrt{a} \geq 0.13\, u^3 + 3u^2.$$

By Lemma 2.2 and $3 \log 2 \geq 2.0794$, the window mass is at most

$$\frac{M}{2}(0.6716\, M + 1) + \frac{M^2}{2.0794} \leq 0.8168\, M^2 + 0.5\, M.$$

Substituting $M \leq 1.5u + 2.42$, the proof closes with the polynomial inequality

$$0.8168\,(1.5u + 2.42)^2 + 0.5\,(1.5u + 2.42) < 0.13\, u^3 + 3u^2 \qquad \text{for } u \geq 6.34,$$

which is checked in Lean. □

2.2. **The zero sum, from the Hasanalizade, Shen and Wong count.** The bound on the zero sum Σ is formalized as `LeanLab.Dudek2014.SigmaH.sigmaBd_of_hsw`, with conclusion `SigmaBd`, which is the bound of Section 2.2 of [Dudek] at $\alpha = 2$, on page 5 of the arXiv version:

Proposition 2.3 (`LeanLab.Dudek2014.SigmaH.sigmaBd_of_hsw`)**.** *Let γ be any family of real ordinates satisfying `HSWN70`, and let Σ be the sum of Dudek's Lemma 5 over $\rho = 1/2 + i\gamma$, written `Zr` in Lean. For every $x \geq 65000$, with $h = \mathrm{hD}(x)$,*

$$|\Sigma| \leq \left(\frac{2xh}{\pi\sqrt{x-h}} + \frac{4h(x+h)^{3/2}}{2\pi x} \right) \log\left(\frac{2x}{h}\right).$$

This theorem takes no Riemann hypothesis. The hypothesis enters only in `theorem2_of_RH_hsw`, through `Zr_eq_Zreal`, which identifies the sum over the family `gZ` with the sum over the nontrivial zeros.

The final bound has the same form as the bound in Dudek's paper. The route differs. From the count `HSWN70`, with its constant 10.2425, the separate head and tail estimates that would give this form do not both hold near $T = 70$: the tail inequality

$$\sum_{|\gamma|>T} \frac{1}{\gamma^2} \leq \frac{\log T}{\pi T}$$

needs about $T \geq 170$, while at the split point $T = 2x/h$ with $x \geq 65000$ we only know $T \geq 70$.

The formal proof keeps the split at $T = 2x/h$ and trades slack between head and tail through four lemmas:

(1) **Head slack,** `head_count`**.** For all $T \geq 70$, the zero count hypothesis `HSWN70` implies

$$\#\{|\gamma| \leq T\} \leq \frac{T\log T}{\pi} - \big(0.89\,T - 22.25 - 0.502\log T\big),$$

so the head count has a positive slack of about $0.89\,T$ below the main term $(T\log T)/\pi$. Indeed, `HSWN70` gives $\#\{|\gamma| \leq T\} \leq 2\,\mathrm{Mt}(T) + 2\,\mathrm{Rt}(T)$, where

$$2\,\mathrm{Mt}(T) = \frac{T\log T}{\pi} - \frac{\log(2\pi)+1}{\pi}\,T + \frac{7}{4}, \qquad \frac{\log(2\pi)+1}{\pi} \geq 0.89.$$

For $T \geq 70$, the lemma `Rt_le` gives $\mathrm{Rt}(T) \leq 0.251\log T + 10.25$, so $2\,\mathrm{Rt}(T) \leq 0.502\log T + 20.5$, and $7/4 + 20.5 = 22.25$. Write $S(T) = 0.89\,T - 22.25 - 0.502\log T$ for this slack.

(2) **Tail with excess,** `tail_inv_sq`**.** For $T \geq 70$, partial summation against the second count in `HSWN70` gives

$$\sum_{|\gamma|>T} \frac{1}{\gamma^2} \leq \frac{\log T}{\pi T} + \frac{E(T)}{T^2}, \qquad E(T) = 1.004\log T + 41.26 - 0.25\,T,$$

and the excess $E(T)$ is positive until about $T = 187$. The lemma `slack` gives $S(T) \geq 0$ and $1.1\,E(T) \leq S(T)$ for $T \geq 70$.

(3) **Comparison of head and tail weights,** `weight_ratio`**.** Let

$$a = h^2(x-h)^{-1/2} \qquad \text{and} \qquad B = 4(x+h)^{3/2}$$

denote the head weight and tail weight, respectively. If $35h \le x$, which is equivalent to $T = 2x/h \ge 70$, then

$$B \le 1.1\, a\, T^2.$$

(4) **Combined split bound,** `split_bound`**.** Let D_i denote the second-difference term

$$D_i = \frac{(x+h)^{\rho_i+1} - 2x^{\rho_i+1} + (x-h)^{\rho_i+1}}{\rho_i(\rho_i+1)}$$

for the i-th zero. Each head term has $\|D_i\| \le a$, by `second_diff_small`, and each tail term has $\|D_i\| \le B/\gamma_i^2$, by `second_diff_large`. The head count and the tail bound give

$$\sum_i \|D_i\| \le \left(\frac{T\log T}{\pi} - S(T)\right) a + B\left(\frac{\log T}{\pi T} + \frac{E(T)}{T^2}\right).$$

Using $B \le 1.1\, a\, T^2$ and $1.1\, E(T) \le S(T)$, the excess from the tail sum is absorbed into the slack of the head count, which yields

$$\sum_i \|D_i\| \le \frac{T\log T}{\pi}\, a + \frac{B\log T}{\pi T}.$$

Since $|\Sigma| \le \sum_i \|D_i\|$, substituting $T = 2x/h$, $a = h^2(x-h)^{-1/2}$, and $B = 4(x+h)^{3/2}$, this upper bound equals the displayed formula in `SigmaBd`.

2.3. **The rest of the formal proof.** The remaining steps of the formalization connect the explicit formula, the positivity of $\mathrm{rhs}(x)$, and the finite prime gap verification to assemble Theorem 2 for all $x \ge 2$:

- **Lemma 4 and Lemma 5,** `lemma4_Zpatch`**.** Lemma 4 of [Dudek] is proved in Lean as `lemma4_Zpatch`, combining the explicit formula `psi1_formula` and the error estimate `error_bound` for all $y > 0$ in the form

$$\left|\psi_1(y) - \left(\frac{y^2}{2} - Z(y) - y\log(2\pi)\right)\right| \le 2.393,$$

which is strictly below the printed bound $12/5 = 2.4$. Here $Z(y)$ is the zero sum for $y > 1$; for $y \le 1$ the Lean definition `Zpatch` takes the value that makes the identity exact, and the proof of Theorem 2 uses Z only for $y > 1$. Lemma 5 of [Dudek] then follows by taking the second difference across $(x-h, x, x+h)$, giving the error bound $48/(5h)$.
- **Positivity for** $x \ge 65000$**,** `rhs_pos`**.** The lemma `rhs_pos` proves by explicit analytic bounds in Lean that $\mathrm{rhs}(x) > 0$ for every real $x \ge 65000$.
- **Large range transfer,** `theorem2_large_of`**.** From the existence of a prime in $(x - \mathrm{hD}(x), x + \mathrm{hD}(x)]$ for all $x \ge 65000$, the theorem `theorem2_large_of` deduces `PrimeInDudekInterval X` for all $X \ge 66799$ via the intermediate value theorem and monotonicity, since $65000 + (2/\pi)\sqrt{65000}\log 65000$ is about 66798.7.
- **Small range verification,** `small_range`**.** The theorem `small_range` proves `PrimeInDudekInterval x` for every real x satisfying $2 \le x < 67001$, with no Riemann hypothesis, by a prime gap scan evaluated by the Lean kernel through `decide +kernel`. The function `gapScan` walks the integers of a tier, keeps the last prime seen, and checks $n < \mathrm{last} + g$ at each n, so every such n has a prime $p \le n$ with $n - p < g$. Table 2 lists the six tiers;

the lemma `gaps_big` is the last one. Since $g \le$ `dudekH`(t) at the start t of each tier and `dudekH` is increasing, the lemma `tier` gives a prime in the interval. Because the small range $[2, 67001)$ and the large range $[66799, \infty)$ overlap, `theorem2_of_RH_hsw` covers all $x \ge 2$.

TABLE 2. The tiers of `small_range`. Each integer n in a tier has a prime $p \le n$ with $n - p < g$.

Integers n	Gap g	`dudekH`(t) at the start t, about
2	1	1.25
3 to 4	2	2.42
5 to 6	2	4.58
7 to 22	4	6.56
23 to 199	14	19.15
200 to 67000	72	95.40

3. Status, cited inputs and data availability

Status of each claim. In Table 3, Kernel means checked by the Lean kernel at the commit below, Text means argued in this note and not in Lean, and Cited means taken from the source and not proved.

TABLE 3. Status of the claims of this note.

Claim	Status	Lean name
Theorem 1.1	Kernel, given `hRH` and `hN`	`theorem2_of_RH_hsw`
Proposition 2.1	Kernel, no hypothesis	`window_holds`
Lemma 2.2	Kernel, no hypothesis	`mass_le`
Proposition 2.3	Kernel, given `HSWN70`	`sigmaBd_of_hsw`
Lemma 4 in the form with 2.393	Kernel, no hypothesis	`lemma4_Zpatch`
Positivity of rhs(x)	Kernel, no hypothesis	`rhs_pos`
Small range	Kernel, no hypothesis	`small_range`
`HSWN70 gZ` follows from Corollary 1.2 of [HSW]	Text	none
Corollary 1.2 of [HSW]	Cited	hypothesis `hN`

Cited inputs. Table 4 lists every cited input of the printed proof and of the formal proof. Corollary 1.2 of [HSW] was read in arXiv v1. The other four sources were not read for this note; they are quoted as [Dudek] states them, by page of its arXiv v2.

Data availability.

- **Registry entry:** Palomar registry `PALOMAR-2026-11-01-000028`, version 1, `https://palomar-registry.org/entry.html?id=PALOMAR-2026-11-01-000028&version=1` [Palomar].
- **Repository:** `https://github.com/sthubbar/goldbach-lean`, directory `entries/dudek2014`, commit `da5cf9edad7896290771dfcab8feee2906399e3d` [Repo].

TABLE 4. Cited inputs. Printed statement, page, how the proof uses it, and Lean form.

Input	Page	Printed statement	Use	Lean form
Corollary 1.2 of [HSW]	arXiv v1, p. 2, display 1.5	$\lvert N(T) - \frac{T}{2\pi}\log\frac{T}{2\pi e}\rvert \le 0.1038\log T + 0.2573\log\log T + 9.3675$ for $T \ge e$	Used for $T \ge 70$, both signs, 7/8 added to Mt and to Rt	`HSWN70 gZ`
Corollary 1 of [Tr]	[Dudek], p. 5	$N(T) < T\log T/(2\pi)$ for $T > 15$	Replaced by `HSWN70`	none
Theorem 14 and 5.5 of [RS]	[Dudek], p. 6	$0.98\sqrt{x} < \psi(x) - \theta(x) < 1.11\sqrt{x} + 3x^{1/3}$ for $x \ge 121$	Replaced by the window bound	`window_holds`
Lemma 1 (ii) of [Sk]	[Dudek], p. 5	Gives the bound on the tail of the zero sum	Replaced from `HSWN70`	`tail_inv_sq`
Value of $\sum_\rho 1/\lvert\rho\rvert^2$, [Da]	[Dudek], p. 4	The value of the sum is explicitly known	Not used	none

- **Toolchain:** Lean and Mathlib version `v4.35.0-rc3` [Mathlib].
- **Build:** in that directory, run `lake exe cache get`, `lake build` and `lake comparator --config=comparator-dudek2014.json`. A fresh public clone of this commit built 9036 jobs in 1196 s of wall time on a 2 core machine.
- **Axioms and comparator check:** The axioms of the compared theorems are solely the standard Lean foundational axioms `propext`, `Quot.sound`, and `Classical.choice`, with `0 sorry`. The comparator passed against the challenge file `ChallengeDudek2014.lean`.
- **Ported files:** sixteen files under `LeanLab/PNTPort/StrongPNT` are from the `PrimeNumberTheoremAnd` project [PNT], commit `c39a751132c88b6e8080b74c74023fd95b3d8be0`, Apache-2.0, credited to its contributors.

Steve Hubbard designed the research program and directed this work. AI tools, Claude and Gemini, were used to write the Lean code and draft the text. AI agents also did the checks of the Lean against the source, and no human expert has reviewed the mathematics. The author takes full responsibility for the content.

REFERENCES

[Dudek] A. W. Dudek, *On the Riemann hypothesis and the difference between primes*, Int. J. Number Theory 11 (2015), no. 3, 771–778, doi:10.1142/S1793042115500426, arXiv:1402.6417.

[HSW] E. Hasanalizade, Q. Shen and P.-J. Wong, *Counting zeros of the Riemann zeta function*, J. Number Theory 235 (2022), 219–241, arXiv:2107.06506.

[CMS] E. Carneiro, M. B. Milinovich and K. Soundararajan, *Fourier optimization and prime gaps*, arXiv:1708.04122v2, first posted 2017, doi:10.4171/CMH/467.

[Da] H. Davenport, *Multiplicative Number Theory*, Springer, 1980.

[DGM] A. W. Dudek, L. Grenié and G. Molteni, *Primes in explicit short intervals on RH*, Int. J. Number Theory 12 (2016), no. 5, 1391–1407, arXiv:1503.05403.

[Sk] S. Skewes, *On the difference* $\pi(x) - \mathrm{li}(x)$ *II*, Proc. London Math. Soc. (3) 5 (1955), 48–70.

[RS] J. B. Rosser and L. Schoenfeld, *Approximate formulas for some functions of prime numbers*, Illinois J. Math. 6, 1962, no. 1, 64–94.

[RaSa] O. Ramaré and Y. Saouter, *Short effective intervals containing primes*, J. Number Theory 98, 2003, no. 1, 10–33.

[Tr] T. S. Trudgian, *An improved upper bound for the argument of the Riemann zeta-function on the critical line*, Math. Comp. 81, 2012, no. 278, 1053–1061.

[Palomar] Palomar registry, entry `PALOMAR-2026-11-01-000028`, version 1, `https://palomar-registry.org/entry.html?id=PALOMAR-2026-11-01-000028&version=1`.

[Repo] S. Hubbard, `goldbach-lean` repository, directory `entries/dudek2014`, commit `da5cf9edad7896290771dfcab8feee2906399e3d`, `https://github.com/sthubbar/goldbach-lean`.

[PNT] `PrimeNumberTheoremAnd`, `https://github.com/AlexKontorovich/PrimeNumberTheoremAnd`, commit `c39a751132c88b6e8080b74c74023fd95b3d8be0`.

[Mathlib] The mathlib Community, *The Lean mathematical library*, `https://github.com/leanprover-community/mathlib4`, version `v4.35.0-rc3`.

Independent Researcher