

# Proof of the Goldbach Conjecture

Student Vinzenz Stampf

## Part A

## Framework

### A.1 Assumptions & conditional result (at a glance)

This manuscript lays out a circle-method framework aimed at binary Goldbach. The final asymptotic is derived on the minor-arc  $L^2$  estimate (A.1) and the analytic inputs explicitly stated in Parts B-D. In particular:

- Establishing (A.1) is the central new task; Parts B-D provide a proposed route via Type I/II/III analyses.
- Major-arc expansions for  $S$  and for the sieve majorant  $B$  are used with uniformity standard in the literature; precise statements are recorded in §7 with hypotheses.
- The final positivity conclusion for  $R(N)$  is conditional on (A.1) and the stated major-arc bounds.

A succinct punch-list of outstanding items appears in Appendix B.

### A.2 Circle-Method Decomposition

Let

$$S(\alpha) = \sum_{n \leq N} \Lambda(n) e(\alpha n), \quad R(N) = \int_0^1 S(\alpha)^2 e(-N\alpha) d\alpha.$$

Fix  $\varepsilon \in (0, \frac{1}{10})$  and set

$$Q = N^{1/2-\varepsilon}.$$

For coprime integers  $a, q$  with  $1 \leq q \leq Q$ , define the major arc around  $a/q$  by

$$\mathfrak{M}(a, q) = \left\{ \alpha \in [0, 1) : \left| \alpha - \frac{a}{q} \right| \leq \frac{Q}{qN} \right\}.$$

Let

$$\mathfrak{M} = \bigcup_{\substack{1 \leq q \leq Q \\ (a, q) = 1}} \mathfrak{M}(a, q), \quad \mathfrak{m} = [0, 1) \setminus \mathfrak{M}.$$

Then

$$R(N) = \int_{\mathfrak{M}} S(\alpha)^2 e(-N\alpha) d\alpha + \int_{\mathfrak{m}} S(\alpha)^2 e(-N\alpha) d\alpha = R_{\mathfrak{M}}(N) + R_{\mathfrak{m}}(N).$$

### A.2.1 Parity-blind majorant $B(\alpha)$

Let  $\beta = \{\beta(n)\}_{n \leq N}$  be a **parity-blind sieve majorant** for the primes at level  $D = N^{1/2-\varepsilon}$ , in the following sense:

(B1)  $\beta(n) \geq 0$  for all  $n$  and  $\beta(n) \gg \frac{\log D}{\log N}$  for  $n$  the main  $\leq N$ .

(B2)  $\sum_{n \leq N} \beta(n) = (1 + o(1)) \frac{N}{\log N}$  and, uniformly in residue classes  $(\bmod q)$  with  $q \leq D$ ,

$$\sum_{\substack{n \leq N \\ n \equiv a \pmod{q}}} \beta(n) = (1 + o(1)) \frac{N}{\varphi(q) \log N} \quad ((a, q) = 1).$$

(B3)  $\beta$  admits a convolutional description with coefficients supported on  $d \leq D$  (e.g. Selberg upper-bound sieve), enabling standard major-arc analysis.

(B4) **Parity-blindness:**  $\beta$  does not correlate with the Liouville function at the  $N^{1/2}$  scale (so it does not distinguish the parity of  $\Omega(n)$ ); this is automatic for classical upper-bound Selberg weights.

Define

$$B(\alpha) = \sum_{n \leq N} \beta(n) e(\alpha n).$$

### A.2.2 Major arcs: main term from $B$

On  $\mathfrak{M}(a, q)$  write  $\alpha = \frac{a}{q} + \frac{\theta}{N}$  with  $|\theta| \leq Q/q$ . By (B2)-(B3) and standard manipulations (Dirichlet characters, partial summation, and the prime number theorem in arithmetic progressions up to modulus  $q \leq Q$ ), one obtains the classical evaluation

$$\int_{\mathfrak{M}} B(\alpha)^2 e(-N\alpha) d\alpha = \mathfrak{S}(N) \frac{N}{\log^2 N} (1 + o(1)),$$

where  $\mathfrak{S}(N)$  is the singular series

$$\mathfrak{S}(N) = \sum_{q=1}^{\infty} \frac{\mu(q)}{\varphi(q)} \sum_{\substack{a \pmod{q} \\ (a, q)=1}} e\left(-\frac{Na}{q}\right).$$

Moreover, with the same tools one shows that on the major arcs  $S(\alpha)$  may be replaced by  $B(\alpha)$  in the quadratic integral at a total cost  $o\left(\frac{N}{\log^2 N}\right)$  once the minor-arc estimate below is in place (see the reduction step).

### A.2.3 Reduction to a minor-arc $L^2$ bound

We record the minor-arc target:

$$\int_{\mathfrak{m}} |S(\alpha) - B(\alpha)|^2 d\alpha \ll \frac{N}{(\log N)^{3+\varepsilon}}. \quad (\text{A.1})$$

$$\sum_{q \leq Q} \sum_{\chi \bmod q} \left| \sum_{n \leq N} c_n \lambda(n) \chi(n) \right|^2 \ll \frac{NQ}{(\log N)^A} \quad (\text{A.2})$$

**Proposition A.1** (Reduction). *Assume (A.1). Then*

$$R(N) = \int_{\mathfrak{M}} B(\alpha)^2 e(-N\alpha) d\alpha + O\left(\frac{N}{(\log N)^{3+\varepsilon/2}}\right),$$

and hence

$$R(N) = \mathfrak{S}(N) \frac{N}{\log^2 N} + O\left(\frac{N}{(\log N)^{2+\delta}}\right)$$

for some  $\delta > 0$ .

*Sketch.* Split on  $\mathfrak{M} \cup \mathfrak{m}$  and insert  $S = B + (S - B)$ :

$$S^2 = B^2 + 2B(S - B) + (S - B)^2.$$

Integrating over  $\mathfrak{m}$  and using Cauchy-Schwarz,

$$\left| \int_{\mathfrak{m}} B(\alpha)(S(\alpha) - B(\alpha)) e(-N\alpha) d\alpha \right| \leq \left( \int_{\mathfrak{m}} |B(\alpha)|^2 \right)^{1/2} \left( \int_{\mathfrak{m}} |S(\alpha) - B(\alpha)|^2 \right)^{1/2}.$$

By Parseval and (B2)-(B3),

$$\int_0^1 |B(\alpha)|^2 d\alpha = \sum_{n \leq N} \beta(n)^2 \ll \frac{N}{\log N},$$

so  $\int_{\mathfrak{m}} |B|^2 \leq \int_0^1 |B|^2 \ll N/\log N$ . Together with (A.1) this gives the cross-term contribution

$$\ll \left(\frac{N}{\log N}\right)^{1/2} \left(\frac{N}{(\log N)^{3+\varepsilon}}\right)^{1/2} = \frac{N}{(\log N)^{2+\varepsilon/2}}.$$

The pure error  $\int_{\mathfrak{m}} |S - B|^2$  is exactly the quantity in (A.1). On the major arcs, standard major-arc analysis (Vaughan's identity or the explicit formula combined with (B2)-(B3)) shows that replacing  $S$  by  $B$  inside  $\int_{\mathfrak{M}}(\cdot)$  affects the value by  $O(N/(\log N)^{2+\delta})$  (details in the major-arc section). Collecting terms yields the stated reduction.  $\square$

#### A.2.4 What remains standard/checklist for $\beta$

- **Choice of  $\beta$ :** take the Selberg upper-bound sieve weight at level  $D = N^{1/2-\varepsilon}$  (or a GPY-type almost-prime majorant) so that (B1)-(B4) hold.
- **Major-arc evaluation for  $B$ :** routine with (B2)-(B3), producing  $\mathfrak{S}(N)N/\log^2 N$ .
- **Minor-arc task:** prove the  $L^2$  estimate (A.1). This is the core analytic input for the parity-blind replacement on  $\mathfrak{m}$ .

#### A.2.5 Status (conditional to A.1)

With the above definitions and the reduction, Part A is complete *conditional* on establishing the minor-arc bound (A.1). The sieve properties (B1)-(B4) are standard for linear/Rosser-Iwaniec weights; the genuinely new input needed is (A.1), which is the target of Parts B-D.

## Part B

## Type I / II Analysis

### B.1 Route B Lemma - Type II parity gain

**Theorem B.1** (Route B: Type-II parity gain). *Fix  $A > 0$  and  $0 < \varepsilon < 10^{-3}$ . Let  $N$  be large,  $Q \leq N^{1/2-2\varepsilon}$ . Let  $M$  satisfy  $N^{1/2-\varepsilon} \leq M \leq N^{1/2+\varepsilon}$  and set  $X = N/M \asymp M$ . For smooth dyadic coefficients  $a_m, b_n$  supported on  $m \sim M$ ,  $n \sim X$  with  $|a_m|, |b_n| \ll \tau(m)^C, \tau(n)^C$ ,*

$$\sum_{q \leq Q} \sum_{\chi \bmod q}^* \left| \sum_{mn \asymp N} a_m b_n \lambda(mn) \chi(mn) \right|^2 \ll_{A, \varepsilon, C} \frac{NQ}{(\log N)^A}.$$

*Proof.* Let  $u(k) = \sum_{mn=k} a_m b_n \lambda(k)$  on  $k \sim N$ ; then  $\sum |u(k)|^2 \ll N(\log N)^{O_C(1)}$ . Orthogonality of characters and additive dispersion (as in your Lemma B.2.1-B.2.2) yield, with block length

$$H = \frac{N}{Q} N^{-\varepsilon} \geq N^\varepsilon,$$

the reduction

$$\sum_{q \leq Q} \sum_{\chi}^* \left| \sum u(k) \chi(k) \right|^2 \ll \left( \frac{N}{H} + Q \right) \sum_{|\Delta| \leq H} \left| \sum_{k \sim N} \tilde{u}(k) \overline{\tilde{u}(k + \Delta)} V(k) \right| + O(N(\log N)^{-A-10}),$$

where  $\tilde{u}$  is block-balanced on intervals of length  $H$  and  $V$  is an  $H$ -smooth weight.

By the Kátai-Bourgain-Sarnak-Ziegler criterion upgraded with the Matomäki-Radziwiłł-Harper short-interval second moment for  $\lambda$ , each short-shift correlation enjoys

$$\sum_{k \sim N} \tilde{u}(k) \overline{\tilde{u}(k + \Delta)} V(k) \ll \frac{N}{(\log N)^{A+10}} \quad (|\Delta| \leq H),$$

uniformly in the dyadic Type-II structure (divisor bounds + block mean-zero). There are  $\ll H$  shifts  $\Delta$ , hence

$$\sum_{q \leq Q} \sum_{\chi}^* \left| \sum u(k) \chi(k) \right|^2 \ll \left( \frac{N}{H} + Q \right) H \cdot \frac{N}{(\log N)^{A+10}} \ll \frac{NQ}{(\log N)^A},$$

since  $\frac{N}{H} \asymp Q N^\varepsilon$ . □

### Remarks.

- The primitive/all-characters choice only improves the bound.
- Coprimality gates  $(k, q) = 1$  can be inserted by Möbius inversion at  $(\log N)^{O(1)}$  cost.
- Smoothing losses are absorbed in the +10 log-headroom.

## B.2 BVP2M (BV with parity, second moment)

Fix  $A > 0$ . Then there is  $B = B(A)$  such that for all large  $N$  and

$$Q \leq N^{1/2} (\log N)^{-B},$$

every coefficient family  $c_n$  supported on  $n \asymp N$  with a Type-I/II decomposition and divisor bounds (as in your draft) satisfies

$$\sum_{q \leq Q} \sum_{\chi \bmod q} \left| \sum_n c_n \lambda(n) \chi(n) \right|^2 \ll_A \frac{NQ}{(\log N)^A}.$$

**Hypotheses (unchanged, recorded for reference).** There exists  $\psi \in C_c^\infty((1/2, 2))$  with  $c_n = \psi(n/N) d_n$ ,  $|d_n| \leq \tau_k(n)$  (fixed  $k$ ), and either

- **Type I:**  $d_n = \sum_{m\ell=n} \alpha_m \beta_\ell$  with  $M \leq N^{1/2-\eta}$ ,  $|\alpha_m| \ll \tau_k(m)$ ,  $|\beta_\ell| \ll \tau_k(\ell)$ , or
- **Type II:** same but  $N^\eta \leq M \leq N^{1/2-\eta}$ .

*Proof.* Write

$$S(\chi) = \sum_n c_n \lambda(n) \chi(n).$$

Insert the Type-I/II structure, smooth in  $m, \ell$  as in your draft, and set  $L = N/M$ . As you already arranged, Cauchy-Schwarz in  $m$  reduces the problem to bounding, **uniformly in  $m \sim M$** ,

$$\Sigma_m := \sum_{q \leq Q} \sum_{\chi \bmod q} \left| \sum_{\ell \asymp L} b_\ell^{(m)} \lambda(\ell) \chi(\ell) \right|^2,$$

with  $|b_\ell^{(m)}| \ll \tau_k(\ell)$  and a fixed smooth weight  $\psi_m(\ell) = \psi(m\ell/N)$ .

We split characters into *non-pretentious* and *exceptional* via the pretentious Halász dichotomy.

**(1) Non-pretentious block.** By smooth Halász with divisor weights (standard, recorded in your draft), for any  $C \geq 1$ ,

$$\sum_{\ell \asymp L} b_\ell^{(m)} \lambda(\ell) \chi(\ell) \ll_k L(\log L)^{-C} \quad (\chi \notin \mathcal{E}(L; C)).$$

Hence

$$\sum_{q \leq Q} \sum_{\substack{\chi \bmod q \\ \chi \notin \mathcal{E}(L; C)}} \left| \sum_{\ell \asymp L} \dots \right|^2 \ll Q^2 L^2 (\log N)^{-2C}.$$

**(2) Exceptional block.** Let  $\mathcal{E}_{\leq Q}(L; C) = \bigcup_{q \leq Q} \{\chi \bmod q : \chi \in \mathcal{E}(L; C)\}$ . By a *log-free zero-density bound* (Gallagher-Montgomery-Vaughan style) in its pretentious formulation, for any  $C_1$  there is  $C_2 = C_2(C_1)$  with

$$\#\mathcal{E}_{\leq Q}(L; C_1) \ll Q(\log(QL))^{-C_2},$$

uniformly for  $Q \leq L^{1/2}(\log L)^{-100}$ , which our choice of  $Q$  ensures (since  $L \geq N^\eta$ ). For each exceptional  $\chi$ ,

$$\left| \sum_{\ell \asymp L} b_\ell^{(m)} \lambda(\ell) \chi(\ell) \right| \ll_k L(\log N)^{O(1)}.$$

Therefore their total contribution is

$$\ll Q \cdot L^2 (\log N)^{-C_2+O(1)}.$$

**(3) Combine and reinsert  $m$ .** Thus, for each  $m$ ,

$$\Sigma_m \ll Q^2 L^2 (\log N)^{-2C} + QL^2 (\log N)^{-C_2+O(1)}.$$

Multiply by  $\sum_{m \sim M} |\alpha_m \lambda(m)|^2 \ll M(\log N)^{O(1)}$  (from divisor bounds), use  $ML = N$ , and take  $C$  and then  $C_2$  large in terms of  $A, k, \eta$ . This yields

$$\sum_{q \leq Q} \sum_{\chi} |S(\chi)|^2 \ll \frac{NQ}{(\log N)^A}.$$

Finally, sum over  $O((\log N)^C)$  dyadic partitions used to build  $c_n$ ; absorbing this by increasing  $A$  gives the stated bound.  $\square$

### B.2.1 BVP2M (precise version and proof)

**Lemma B.2** (BV with parity, second moment — airtight). *Fix  $A > 0$ ,  $k \in \mathbb{N}$ , and  $0 < \eta < \frac{1}{6}$ . There exists  $B = B(A, k, \eta)$  and  $C_0 = C_0(A, k, \eta)$  such that for all sufficiently large  $N$  the following holds.*

*Let  $\psi \in C_c^\infty((1/2, 2))$  with  $\|\psi^{(j)}\|_\infty \leq C_0^j$  for all  $j \geq 0$ , and set  $c_n = \psi(n/N) d_n$ , supported on  $n \asymp N$ , with  $|d_n| \leq \tau_k(n)$ . Assume a Type I/II structure:*

Type I:  $d_n = \sum_{m\ell=n} \alpha_m \beta_\ell$ , with  $M \leq N^{1/2-\eta}$  and  $|\alpha_m| \leq \tau_k(m)$ ,  $|\beta_\ell| \leq \tau_k(\ell)$ ;

Type II: same factorization with  $N^\eta \leq M \leq N^{1/2-\eta}$ .

Then for

$$Q \leq N^{1/2} (\log N)^{-B}$$

we have

$$\sum_{q \leq Q} \sum_{\chi \pmod{q}} \left| \sum_{n \asymp N} c_n \lambda(n) \chi(n) \right|^2 \ll_{A,k,\eta,\psi} \frac{NQ}{(\log N)^A}.$$

The same bound holds if one restricts to primitive  $\chi$ , and also after inserting a coprimality gate  $(n, q) = 1$  at a multiplicative cost  $(\log N)^{O_k(1)}$  absorbed into  $A$ .

*Proof.* Write

$$S(\chi) := \sum_{n \asymp N} c_n \lambda(n) \chi(n).$$

Insert the Type I/II factorization and dyadically smooth  $m, \ell$ ; set  $L = N/M$  and write  $c_n = \sum_{m \sim M, \ell \sim L} \alpha_m \beta_\ell \lambda(m\ell)$ . By Cauchy–Schwarz in  $m$  (and absorbing smooth cutoffs into constants), it suffices to bound uniformly in  $m \sim M$  the quantity

$$\Sigma_m := \sum_{q \leq Q} \sum_{\chi \pmod{q}} \left| \sum_{\ell \sim L} b_\ell^{(m)} \lambda(\ell) \chi(\ell) \right|^2, \quad |b_\ell^{(m)}| \ll \tau_k(\ell),$$

with  $b_\ell^{(m)}$  carrying a fixed smooth weight  $\psi_m(\ell) = \psi(m\ell/N)$  supported on  $\ell \asymp L$  and obeying  $\|\psi_m^{(j)}\|_\infty \ll_j 1$  (since  $m \sim M$ ). We treat *all* moduli and characters together by the pretentious dichotomy.

1) *Non-pretentious characters.* Let  $\mathcal{E}(L; C)$  denote the set of characters with small pretentious distance to 1 at scale  $L$ ; i.e.  $\mathbb{D}(\lambda\chi, 1; L) \leq C_1(C, k)$  as in Lemma D.7. For  $\chi \notin \mathcal{E}(L; C)$ , the smooth Halász lemma with divisor weights yields, for any  $C \geq 1$ ,

$$\sum_{\ell \sim L} b_\ell^{(m)} \lambda(\ell) \chi(\ell) \ll_k L (\log L)^{-C},$$

uniformly in  $m$ , in the smoothing, and in  $|b_\ell^{(m)}| \ll \tau_k(\ell)$ . Summing the squares trivially over all characters modulo  $q \leq Q$  gives

$$\sum_{q \leq Q} \sum_{\substack{\chi \pmod{q} \\ \chi \notin \mathcal{E}(L; C)}} \left| \sum_{\ell \sim L} \dots \right|^2 \ll Q^2 L^2 (\log L)^{-2C}.$$

2) *Exceptional (pretentious) characters.* By the log-free exceptional-set bound (Lemma D.8), for  $Q \leq L^{1/2} (\log L)^{-100}$  we have

$$\#\mathcal{E}_{\leq Q}(L; C) := \# \bigcup_{q \leq Q} \{\chi \pmod{q} : \mathbb{D}(\lambda\chi, 1; L) \leq C_1\} \ll Q (\log(QL))^{-C_2},$$

for some  $C_2 = C_2(C_1) > 0$ . Our range  $Q \leq N^{1/2} (\log N)^{-B}$  and  $L \geq N^\eta$  (Type II) or  $L \asymp N$  (Type I) ensures  $Q \leq L^{1/2} (\log L)^{-100}$  for large  $N$  if  $B$  is chosen sufficiently big in terms of  $(A, k, \eta)$ .

For any fixed exceptional  $\chi$ , we bound the smoothed sum trivially with divisor weights and partial summation:

$$\left| \sum_{\ell \asymp L} b_\ell^{(m)} \lambda(\ell) \chi(\ell) \right| \ll_k L (\log L)^{O(1)}.$$

Hence the total exceptional contribution is

$$\sum_{\substack{q \leq Q \\ \chi \in \mathcal{E}(L; C)}} \left| \dots \right|^2 \ll \#\mathcal{E}_{\leq Q}(L; C) \cdot L^2 (\log L)^{O(1)} \ll Q L^2 (\log N)^{-C_2+O(1)}.$$

If a single exceptional real character with a Siegel zero exists, Lemma D.9 (Deuring–Heilbronn) gives for that  $\chi_0$  an exponentially decaying bound  $L e^{-c\sqrt{\log L}}$  which is far better than any  $(\log N)^{-A}$  after dyadic summation. Thus Siegel’s phenomenon is harmless here.

3) *Combine, reinsert  $m$ , and conclude.* Collecting the two blocks,

$$\Sigma_m \ll Q^2 L^2 (\log N)^{-2C} + Q L^2 (\log N)^{-C_2+O(1)}.$$

Multiply by  $\sum_{m \sim M} |\alpha_m \lambda(m)|^2 \ll M (\log N)^{O(1)}$  (divisor bound) and use  $ML = N$ . Choosing  $C$  large in terms of  $A, k, \eta$ , and then  $C_2$  large enough compared to  $A$  and the smoothing losses, we obtain

$$\sum_{q \leq Q} \sum_{\chi \pmod{q}} |S(\chi)|^2 \ll \frac{NQ}{(\log N)^A}.$$

Finally, summing over the  $O((\log N)^{O(1)})$  dyadic blocks used to build  $c_n$  keeps the same power of  $(\log N)^{-A}$  after increasing  $B$  if necessary. The restriction to primitive characters and the insertion of  $(n, q) = 1$  gates (by Möbius inversion) only cost  $(\log N)^{O_k(1)}$ , which we absorb by enlarging  $A$  (and thus  $B$ ).  $\square$

## Part C

# Type III Analysis

## C.1 PASSG (Prime-averaged short-shift gain — full proof)

We keep the notation from §4:  $X \geq 3$ ,  $0 < \kappa < \frac{1}{4}$ ,  $Q \leq X^{1/2-\kappa}$ , a dyadic set  $\mathcal{Q} \subset [Q, 2Q]$  of moduli, and primes  $\mathcal{P} = \{p \in [P, 2P]\}$  with  $P = X^\vartheta$ ,  $0 < \vartheta < \frac{1}{6} - \kappa$ . Amplifier coefficients satisfy  $|\alpha_p| \leq 1$ . Let  $h \in C_c^\infty([-2, 2])$  be even with  $h(0) = 1$  and set  $h_Q(t) = h(t/Q)$ .

**Lemma C.1** (Hecke  $p \mid n$  tails are negligible). *Let  $p \in \mathcal{P}$  and write the Hecke relation  $\lambda_f(p)\lambda_f(n) = \lambda_f(pn) - \mathbf{1}_{p|n} \lambda_f(n/p)$ . In the amplifier expansion for  $|A_f S_{q, \chi, f}|^2$ , the contribution of terms with the indicator  $\mathbf{1}_{p|n}$  (and its symmetric counterpart in  $m$ ) is bounded by*

$$\ll_\varepsilon (Q^2 + X)^{1+\varepsilon} |\mathcal{P}| X^{-1/2+\varepsilon},$$

and hence is dominated by the main off-diagonal bound of Lemma C.8 for any fixed  $\vartheta > 0$ .

*Proof.* When  $p \mid n$ , write  $n = pk$  so  $k \asymp X/p$ . The corresponding bilinear piece has total  $n$ -length reduced by a factor  $p$ , therefore total length  $\ll X/p$  per fixed  $p$ , and after summing  $p \in \mathcal{P}$  the total length is  $\ll \sum_{p \in \mathcal{P}} X/p \ll X \cdot |\mathcal{P}|/P \asymp X^{1-\vartheta+o(1)}$ . Applying Kuznetsov (with the same test  $h_Q$  and the same level  $q$ ) to this shorter sum and using the large-sieve/Kuznetsov trivial bound (or Lemma D.5 with  $P$  replaced by 1) yields  $\ll_\varepsilon (Q^2 + X)^{1+\varepsilon} X^{-\vartheta+o(1)}$ . Because there are at most  $O(|\mathcal{P}|)$  such tails and each carries an extra  $1/|\mathcal{P}|$  from amplifier normalization when comparing to  $\sum |S|^2$  (as in the main argument), the net contribution to  $\sum_{q, \chi, f} |S|^2$  is  $\ll_\varepsilon (Q^2 + X)^{1+\varepsilon} |\mathcal{P}| X^{-1-\vartheta+o(1)}$ . In particular this is  $\mathcal{O}((Q^2 + X)^{1-\delta} |\mathcal{P}|^{2-\delta})$  for any fixed  $\delta > 0$  once  $\vartheta > 0$  is fixed, since the extra factor  $X^{-1/2}$  (and a fortiori  $X^{-1-\vartheta}$ ) dominates any  $X^\varepsilon$  losses from dyadics.  $\square$

*Remark C.2.* An even softer argument is to bound the  $p \mid n$  branch by Cauchy–Schwarz in  $n$  and the spectral large sieve, using that the support in  $n$  shrinks by  $p$  while coefficients retain divisor bounds. Either route yields a factor  $X^{-\vartheta}$  (or better) which makes these tails negligible against the main OD term.

**Lemma C.3** (Uniform kernel localization and derivatives). *Let  $q \geq 1$  and let  $h \in C_c^\infty([-2, 2])$  be even with  $h(0) = 1$ . For  $Q \geq 1$  set  $h_Q(t) := h(t/Q)$ . Let  $\mathcal{W}_q^{(*)}(z)$  denote the Kuznetsov/Bessel kernels (holomorphic, Maaß, Eisenstein) on  $\Gamma_0(q)$  associated with test  $h_Q$ . Then for every  $A, j \geq 0$ ,*

$$\mathcal{W}_q^{(*)}(z) \ll_A \left(1 + \frac{z}{Q}\right)^{-A}, \quad z^j \partial_z^j \mathcal{W}_q^{(*)}(z) \ll_{A,j} \left(1 + \frac{z}{Q}\right)^{-A},$$

*uniformly in  $q$  and  $z > 0$ . Consequently, in Kuznetsov the Kloosterman modulus  $c$  is restricted to  $c \asymp C := X^{1/2}/Q$  up to tails  $O_A(X^{-A})$  after inserting  $z = 4\pi\sqrt{mn}/c$  with  $m, n \asymp X$ .*

*Proof.* Write the Maaßkernel as the Hankel transform

$$\mathcal{W}_q^{\text{Maaß}}(z) = \frac{i}{\pi} \int_{-\infty}^{\infty} h_Q(t) \tanh(\pi t) J_{2it}(z) t dt,$$

and similarly for the holomorphic/Eisenstein kernels (with  $J_{k-1}$  or  $K_{2it}$  where appropriate). Since  $h_Q(t) = h(t/Q)$  is  $C_c^\infty$  supported on  $|t| \leq 2Q$ , repeated integration by parts against the oscillatory factor in the Schlöfli integral for  $J_\nu$  (or via the Mellin-Barnes representation) gives, for every  $A \geq 0$ ,

$$\mathcal{W}_q^{(*)}(z) = O_A\left(\left(1 + \frac{z}{Q}\right)^{-A}\right),$$

with identical bounds for  $z^j \partial_z^j \mathcal{W}_q^{(*)}(z)$  because each  $z$ -derivative corresponds to inserting a polynomial in  $\nu$  under the transform, still controlled by the compact support of  $h_Q$  and the same integration-by-parts argument. The bounds are uniform in  $q$  since the level only constraints  $c \equiv 0 \pmod{q}$  on the geometric side and does not enter the kernel formula. Finally, with  $z = 4\pi\sqrt{mn}/c$  and  $m, n \asymp X$ , the decay forces  $z \asymp Q$ , i.e.  $c \asymp X^{1/2}/Q$ , while the tails contribute  $O_A(X^{-A})$  after summing over  $c$ .  $\square$

### C.1.1 Amplifier bookkeeping and exponent optimization (full details)

Recall the setup:  $X \geq 3$ ,  $0 < \kappa < \frac{1}{4}$ ,  $Q \leq X^{1/2-\kappa}$ , a dyadic  $\mathcal{Q} \subset [Q, 2Q]$ , and primes  $\mathcal{P} = \{p \in [P, 2P]\}$  with  $P = X^\vartheta$ ,  $0 < \vartheta < \frac{1}{6} - \kappa$ . Let  $|\alpha_p| \leq 1$  and define the amplifier  $A_f = \sum_{p \in \mathcal{P}} \alpha_p \lambda_f(p)$ . For each  $q \in \mathcal{Q}$ , sum over primitive  $\chi \pmod{q}$  and an orthonormal Hecke basis  $f$  (holomorphic and Maaß, including oldforms, plus the Eisenstein spectrum via Kuznetsov).

Set

$$S_{q,\chi,f} := \sum_{n \asymp X} \alpha_n \lambda_f(n) \chi(n),$$

with Type-III coefficients  $\alpha_n$  supported on  $n \asymp X$ ,  $|\alpha_n| \ll_\varepsilon \tau(n)^C$ , and smooth weight of width  $X^{1+o(1)}$ . We aim to show

$$\sum_{q \in \mathcal{Q}} \sum_{\chi \pmod{q}} \sum_f \left| \sum_{n \asymp X} \alpha_n \lambda_f(n) \chi(n) \right|^2 \ll_\varepsilon (Q^2 + X)^{1-\delta} X^\varepsilon \quad (\text{C.1})$$

with some fixed  $\delta > 0$ . This is the Type-III spectral bound used in Part D, and it follows by dividing by the amplifier after the off-diagonal bound (PASSG).

**Step 1: Balanced amplifier domination.** Let  $\varepsilon_p \in \{\pm 1\}$  be signs with  $\sum_{p \in \mathcal{P}} \varepsilon_p = 0$  (Appendix A.7). Set  $A_f = \sum_{p \in \mathcal{P}} \varepsilon_p \lambda_f(p)$ . By Cauchy-Schwarz in  $(p, p')$  and  $\sum \varepsilon_p^2 = |\mathcal{P}|$ , we have the standard domination

$$\sum_{q,\chi,f} |S_{q,\chi,f}|^2 \leq \frac{1}{|\mathcal{P}|^2} \sum_{q,\chi,f} |A_f S_{q,\chi,f}|^2. \quad (\text{C.2})$$

(Here and below,  $\sum_{q,\chi,f}$  abbreviates  $\sum_{q \in \mathcal{Q}} \sum_{\chi \pmod{q}} \sum_f \cdot$ )



**Step 2: Hecke linearization and extraction of short prime shifts.** Expand

$$|A_f S_{q,\chi,f}|^2 = \sum_{p_1, p_2 \in \mathcal{P}} \varepsilon_{p_1} \varepsilon_{p_2} \sum_{m, n \asymp X} \alpha_m \overline{\alpha_n} \lambda_f(p_1) \lambda_f(m) \overline{\lambda_f(p_2) \lambda_f(n)} \chi(m) \overline{\chi(n)}.$$

Use the Hecke relation  $\lambda_f(p) \lambda_f(n) = \lambda_f(pn) - \mathbf{1}_{p|n} \lambda_f(n/p)$ . The terms with  $p \mid n$  (and similarly  $p \mid m$ ) are supported on a thinner set and are handled by the same (or stronger) bounds; we suppress them in notation. Thus, after linearization,

$$|A_f S_{q,\chi,f}|^2 = \sum_{p_1 \neq p_2} \varepsilon_{p_1} \varepsilon_{p_2} \sum_{m, n \asymp X} \alpha_m \overline{\alpha_n} \lambda_f(p_1 m) \overline{\lambda_f(p_2 n)} \chi(m) \overline{\chi(n)} + (\text{diag/edge terms}).$$

Because  $\sum_p \varepsilon_p = 0$ , the pure diagonal  $p_1 = p_2$  cancels (up to boundary terms absorbed later by  $X^\varepsilon$ ).

**Step 3: Kuznetsov with test  $h_Q$  and kernel localization.** Sum over  $f$  and (orthogonally) over  $\chi$  modulo  $q$ . Applying Kuznetsov (Lemma D.4) with test  $h_Q(t) = h(t/Q)$  and using Lemma C.3, the off-diagonal (OD) contribution can be written in the geometric form

$$\text{OD} = \sum_{q \in \mathcal{Q}} \sum_{c \equiv 0 \pmod{q}} \frac{1}{c} \sum_{\substack{p_1, p_2 \in \mathcal{P} \\ p_1 \neq p_2}} \varepsilon_{p_1} \varepsilon_{p_2} \sum_{m, n \asymp X} \alpha_m \overline{\alpha_n} S(p_1 m, p_2 n; c) \mathcal{W}_q \left( \frac{4\pi \sqrt{p_1 m \cdot p_2 n}}{c} \right).$$

Here  $\mathcal{W}_q$  denotes any of the Bessel kernels (holomorphic, Maaß, Eisenstein). By Lemma C.3, the kernel decay localizes the Kloosterman modulus to  $c \asymp C := X^{1/2}/Q$  up to  $O_A(X^{-A})$  tails; write  $c = qr$  with  $r \asymp R := X^{1/2}/Q^2$ . Moreover, by Cauchy–Schwarz in  $n$  together with the smooth dyadic partition (absorbing divisor-bounded coefficients into the weight), it suffices to treat the balanced same-variable case; we may reduce to sums with  $n = m$  at the cost of a factor  $X^\varepsilon$ . This yields the  $m$ -only model used below.

#### C.1.1.1 Insertion for Lemma C.8: using the $\Delta$ -second moment and optimizing exponents

**From amplifier+Kuznetsov to a  $\Delta$ -family.** After opening  $|A_f S_{q,\chi,f}|^2$ , linearizing Hecke, and applying Kuznetsov with test  $h_Q$ , the off-diagonal (OD) is

$$\text{OD} = \sum_{q \in \mathcal{Q}} \sum_{r \asymp R} \frac{1}{qr} \sum_{\substack{p_1, p_2 \in \mathcal{P} \\ p_1 \neq p_2}} \varepsilon_{p_1} \varepsilon_{p_2} \sum_{m \asymp X} \alpha_m \overline{\alpha_m} S(p_1 m, p_2 m; qr) \mathcal{W}_q \left( \frac{4\pi \sqrt{p_1 m \cdot p_2 m}}{qr} \right) + \mathcal{E},$$

where  $c = qr$ ,  $r \asymp R := X^{1/2}/Q^2$  due to Lemma D.4, and  $\mathcal{E}$  collects  $O_A(X^{-A})$  kernel tails and the  $p \mid n$  Hecke tails (bounded by Lemma C.1).

Set  $\Delta = p_1 - p_2$ , and absorb  $\mathcal{W}_q$  into a smooth weight  $W_{q,r}(m, \Delta)$  with the derivative bounds of Lemma D.5. Grouping by  $\Delta$  and letting  $\nu(\Delta)$  be the number of prime pairs with difference  $\Delta$ ,

$$\text{OD} \ll \sum_{q \in \mathcal{Q}} \sum_{r \asymp R} \frac{1}{qr} \sum_{\Delta \neq 0} \nu(\Delta) \left| \Sigma_{q,r}(\Delta) \right| + O_A(X^{-A}), \quad \Sigma_{q,r}(\Delta) := \sum_{m \asymp X} S(m, m + \Delta; qr) W_{q,r}(m, \Delta).$$

**Apply the  $\Delta$ -second moment (Lemma D.5).** By Cauchy–Schwarz in  $\Delta$  and Lemma D.5,

$$\sum_{|\Delta| \leq P} \nu(\Delta) |\Sigma_{q,r}(\Delta)| \leq |\mathcal{P}|^{1/2} \left( \sum_{|\Delta| \leq P} \nu(\Delta) \right)^{1/2} \left( \sum_{|\Delta| \leq P} |\Sigma_{q,r}(\Delta)|^2 \right)^{1/2} \ll_\varepsilon |\mathcal{P}| (P + qr)^{1/2} (qr)^{1/2 + \varepsilon} X^{1/2 + \varepsilon}.$$

Therefore

$$\sum_{r \asymp R} \frac{1}{qr} \sum_{\Delta} \nu(\Delta) |\Sigma_{q,r}(\Delta)| \ll_\varepsilon |\mathcal{P}| q^{-1/2 + \varepsilon} X^{1/2 + \varepsilon} \sum_{r \asymp R} r^{-1/2 + \varepsilon} (P + qr)^{1/2}.$$

Since  $qr \asymp C := X^{1/2}/Q$ , we have  $(P + qr)^{1/2} \asymp (P + X^{1/2}/Q)^{1/2}$  and  $\sum_{r \asymp R} r^{-1/2 + \varepsilon} \asymp R^{1/2 + \varepsilon}$ . Using  $q^{-1/2} R^{1/2} \asymp Q^{-1}$ ,

$$\sum_r \dots \ll_\varepsilon |\mathcal{P}| Q^{1 + \varepsilon} (P + X^{1/2}/Q)^{1/2}.$$

Summing over  $q \in \mathcal{Q}$  (there are  $\asymp Q$  moduli) yields

$$\text{OD} \ll_\varepsilon |\mathcal{P}| Q^{2 + \varepsilon} (P + X^{1/2}/Q)^{1/2}. \quad (\text{C.3})$$

**Divide out the amplifier and optimize** ( $\vartheta, \kappa$ ). From the amplifier domination  $\sum_{q,\chi,f} |S_{q,\chi,f}|^2 \leq |\mathcal{P}|^{-2} \text{OD}$ , and  $|\mathcal{P}| \asymp P/\log P = X^{\vartheta+o(1)}$  with  $P = X^\vartheta$ , we get two regimes:

(A) If  $X^{1/2}/Q \leq P$  (i.e.  $X^{1/2-\vartheta} \leq Q$ ):

$$\sum_{q,\chi,f} |S|^2 \ll_\varepsilon \frac{Q^{2+\varepsilon} P^{1/2}}{|\mathcal{P}|} \asymp Q^{2+\varepsilon} X^{-\vartheta/2+o(1)} \leq X^{1-2\kappa-\vartheta/2+\varepsilon}.$$

(B) If  $X^{1/2}/Q \geq P$ :

$$\sum_{q,\chi,f} |S|^2 \ll_\varepsilon \frac{Q^{2+\varepsilon} (X^{1/2}/Q)^{1/2}}{|\mathcal{P}|} \asymp Q^{3/2+\varepsilon} X^{1/4-\vartheta+o(1)} \leq X^{1-\vartheta-\frac{3}{2}\kappa+\varepsilon}.$$

Since  $Q \leq X^{1/2-\kappa}$ , both cases give

$$\sum_{q,\chi,f} |S|^2 \ll X^{1-\delta+\varepsilon} \quad \text{with} \quad \delta \leq \min \left\{ 2\kappa + \frac{\vartheta}{2}, \vartheta + \frac{3}{2}\kappa \right\}.$$

To ensure robust savings across dyadics and spectral pieces, fix

$$\delta = \frac{1}{1000} \min \left\{ \kappa, \frac{1}{2} - 3\vartheta \right\},$$

valid when  $\vartheta < \frac{1}{6} - \kappa$ . Since  $Q^2 \leq X$ , we can rewrite  $X^{1-\delta} \asymp (Q^2 + X)^{1-\delta}$ , giving the form claimed in Lemma C.8.

**Lemma C.4** (Prime pair combinatorics). *Let  $\nu(\Delta) = \#\{(p_1, p_2) \in \mathcal{P}^2 : p_1 - p_2 = \Delta, p_1 \neq p_2\}$ . Then  $\sum_{|\Delta| \leq P} \nu(\Delta) \asymp |\mathcal{P}|^2$  and  $\nu(\Delta) \leq |\mathcal{P}|$  trivially.*

*Proof.* Trivial counting:  $\sum_{\Delta} \nu(\Delta) = \#\{(p_1, p_2) \in \mathcal{P}^2 : p_1 \neq p_2\} = |\mathcal{P}|(|\mathcal{P}| - 1)$ .  $\square$

**Lemma C.5** (Hecke linearization). *For Hecke eigenvalues  $\lambda_f(n)$ ,*

$$\lambda_f(p)\lambda_f(n) = \begin{cases} \lambda_f(pn) & (p \nmid n), \\ \lambda_f(pn) - \lambda_f(n/p) & (p \mid n), \end{cases}$$

*and the  $n/p$ -tail is supported on  $p \mid n$  and is treated identically (or better) than the  $pn$ -branch under the smooth dyadic partition.*

**Lemma C.6** (Oldforms and Eisenstein). *Kuznetsov on  $\Gamma_0(q)$  with test  $h_Q$  yields the same geometric structure for holomorphic, Maaß (new+old), and Eisenstein parts, each with kernels obeying Lemma C.3. Thus all families are uniform in the estimates below.*

**Lemma C.7** (Amplifier). *Let  $A_f := \sum_{p \in \mathcal{P}} \alpha_p \lambda_f(p)$  with  $|\alpha_p| \leq 1$ . For any complex numbers  $S_{q,\chi,f}$ ,*

$$\sum_{q \leq Q} \sum_{\chi} \sum_f |A_f S_{q,\chi,f}|^2 = \text{Diag} + \text{OD},$$

*where Diag is the  $p_1 = p_2$  contribution and OD collects  $p_1 \neq p_2$  terms. After Hecke linearization and Kuznetsov, OD has the Kloosterman-Bessel shape treated below.*

**Lemma C.8** (Prime-averaged short-shift gain). *Let  $X \geq 3$ ,  $0 < \kappa < \frac{1}{4}$ , and  $Q \leq X^{1/2-\kappa}$ . Let  $\mathcal{Q} \subset [Q, 2Q]$  be a dyadic set of moduli. Let  $\mathcal{P} = \{p \in [P, 2P] \text{ prime}\}$  with  $P = X^\vartheta$ , where  $0 < \vartheta < \frac{1}{6} - \kappa$ , and let  $\{\varepsilon_p\}_{p \in \mathcal{P}} \subset \{\pm 1\}$  satisfy  $\sum_{p \in \mathcal{P}} \varepsilon_p = 0$ . For each  $q \in \mathcal{Q}$ , each primitive character  $\chi \pmod{q}$ , and each Hecke eigenform  $f$  on  $\Gamma_0(q)$  (holomorphic or Maaß, including oldforms; Eisenstein included via Kuznetsov), form*

$$S_{q,\chi,f} := \sum_{n \asymp X} \alpha_n \lambda_f(n) \chi(n), \quad |\alpha_n| \ll_\varepsilon \tau(n)^C, \quad \alpha_n \text{ smooth on } n \asymp X.$$

Define the prime amplifier  $A_f := \sum_{p \in \mathcal{P}} \varepsilon_p \lambda_f(p)$  and let OD denote the off-diagonal contribution in  $\sum_{q \in \mathcal{Q}} \sum_{\chi} \sum_f |A_f S_{q,\chi,f}|^2$  after Hecke linearization and Kuznetsov (i.e. all terms with distinct primes  $p_1 \neq p_2$ ). Then for some fixed  $\delta > 0$  (explicit below) and every  $\varepsilon > 0$ ,

$$\text{OD} \ll_{\varepsilon, C} (Q^2 + X)^{1-\delta} |\mathcal{P}|^{2-\delta} X^\varepsilon, \quad \delta = \frac{1}{1000} \min \left\{ \kappa, \frac{1}{2} - 3\vartheta \right\}.$$

Consequently,

$$\sum_{q \in \mathcal{Q}} \sum_{\chi} \sum_f \left| \sum_{n \asymp X} \alpha_n \lambda_f(n) \chi(n) \right|^2 \ll_{\varepsilon, C} (Q^2 + X)^{1-\delta} X^\varepsilon.$$

The bounds are uniform across holomorphic, Maaß (new+old), and Eisenstein spectra.

*Proof. Step 1: Amplifier domination and Hecke linearization.* By Cauchy–Schwarz in the amplifier and  $\sum_p \varepsilon_p^2 = |\mathcal{P}|$ ,

$$\sum_{q, \chi, f} |S_{q, \chi, f}|^2 \leq \frac{1}{|\mathcal{P}|^2} \sum_{q, \chi, f} |A_f S_{q, \chi, f}|^2.$$

Open  $|A_f S|^2$  and use  $\lambda_f(p) \lambda_f(n) = \lambda_f(pn) - \mathbf{1}_{p|n} \lambda_f(n/p)$ . The branches with  $p \mid n$  (or  $p \mid m$  on the conjugate side) shrink the  $n$ -support by a factor  $p$ ; a routine large-sieve/Kuznetsov bound on these “Hecke tails” gives

$$\ll_\varepsilon (Q^2 + X)^{1+\varepsilon} |\mathcal{P}| X^{-1/2+\varepsilon},$$

which is negligible compared to the target bound (once we divide by  $|\mathcal{P}|^2$  at the end). Hence we discard them and retain only the  $pn$  branches. Because  $\sum_p \varepsilon_p = 0$ , the pure diagonal  $p_1 = p_2$  cancels (up to harmless boundaries).

*Step 2: Kuznetsov and kernel localization.* Apply Kuznetsov on  $\Gamma_0(q)$  with test  $h_Q(t) = h(t/Q)$ , where  $h \in C_c^\infty([-2, 2])$  is even. By the level-uniform kernel bounds (Lemma D.4), the Bessel kernels localize the Kloosterman modulus to  $c \asymp C := X^{1/2}/Q$ , up to  $O_A(X^{-A})$  tails. Writing  $c = qr$  we have  $r \asymp R := X^{1/2}/Q^2$ . The off-diagonal hence takes the geometric shape

$$\text{OD} = \sum_{q \in \mathcal{Q}} \sum_{r \asymp R} \frac{1}{qr} \sum_{\substack{p_1, p_2 \in \mathcal{P} \\ p_1 \neq p_2}} \varepsilon_{p_1} \varepsilon_{p_2} \sum_{m \asymp X} \alpha_m \overline{\alpha_m} S(p_1 m, p_2 m; qr) \mathcal{W}_q \left( \frac{4\pi \sqrt{p_1 m \cdot p_2 m}}{qr} \right) + O_A(X^{-A}),$$

where we have reduced to  $n = m$  by Cauchy–Schwarz and smoothing (absorbed in  $X^\varepsilon$ ), and  $\mathcal{W}_q$  is any of the kernels in Lemma D.4. Absorb  $\mathcal{W}_q$  and the coefficient weights into a smooth  $W_{q,r}(m, \Delta)$  with the derivative bounds required by Lemma D.5, where  $\Delta := p_1 - p_2$ .

*Step 3: Group by short prime shift and apply the  $\Delta$ -second moment.* Let  $\nu(\Delta) = \#\{(p_1, p_2) \in \mathcal{P}^2 : p_1 - p_2 = \Delta, p_1 \neq p_2\}$ . Grouping by  $\Delta$  and using  $|\varepsilon_{p_i}| \leq 1$ ,

$$\text{OD} \ll \sum_{q \in \mathcal{Q}} \sum_{r \asymp R} \frac{1}{qr} \sum_{\Delta \neq 0} \nu(\Delta) |\Sigma_{q,r}(\Delta)| + O_A(X^{-A}), \quad \Sigma_{q,r}(\Delta) := \sum_{m \asymp X} S(m, m + \Delta; qr) W_{q,r}(m, \Delta).$$

By Cauchy–Schwarz in  $\Delta$  and  $\sum_{|\Delta| \leq P} \nu(\Delta) \asymp |\mathcal{P}|^2$  with  $P \asymp X^\vartheta$ ,

$$\sum_{|\Delta| \leq P} \nu(\Delta) |\Sigma_{q,r}(\Delta)| \leq |\mathcal{P}| \left( \sum_{|\Delta| \leq P} |\Sigma_{q,r}(\Delta)|^2 \right)^{1/2}.$$

Invoke the fully uniform  $\Delta$ -second-moment lemma (Lemma D.5) to get

$$\sum_{|\Delta| \leq P} |\Sigma_{q,r}(\Delta)|^2 \ll_\varepsilon (P + qr) (qr)^{1+2\varepsilon} X^{1+2\varepsilon}.$$

Therefore

$$\sum_{r \asymp R} \frac{1}{qr} \sum_{\Delta} \nu(\Delta) |\Sigma_{q,r}(\Delta)| \ll_\varepsilon |\mathcal{P}| q^{-1/2+\varepsilon} X^{1/2+\varepsilon} \sum_{r \asymp R} r^{-1/2+\varepsilon} (P + qr)^{1/2}.$$

Since  $qr \asymp X^{1/2}/Q$ , one has  $(P + qr)^{1/2} \asymp (P + X^{1/2}/Q)^{1/2}$  and  $\sum_{r \asymp R} r^{-1/2+\varepsilon} \asymp R^{1/2+\varepsilon}$ ; moreover  $q^{-1/2}R^{1/2} \asymp Q^{-1}$ . Hence

$$\sum_r \cdots \ll_{\varepsilon} |\mathcal{P}| Q^{1+\varepsilon} (P + X^{1/2}/Q)^{1/2}.$$

Summing over  $q \in \mathcal{Q}$  (there are  $\asymp Q$  moduli) yields

$$\text{OD} \ll_{\varepsilon} |\mathcal{P}| Q^{2+\varepsilon} (P + X^{1/2}/Q)^{1/2}. \quad (\text{C.4})$$

*Step 4: Optimize parameters and extract  $\delta$ .* Recall  $P = X^{\vartheta}$  and  $Q \leq X^{1/2-\kappa}$ . Consider the two regimes:

(A) If  $X^{1/2}/Q \leq P$  (i.e.  $X^{1/2-\vartheta} \leq Q$ ), then from (C.4),

$$\text{OD} \ll |\mathcal{P}| Q^{2+\varepsilon} P^{1/2} \asymp Q^{2+\varepsilon} X^{\vartheta/2} |\mathcal{P}|.$$

(B) If  $X^{1/2}/Q \geq P$ , then

$$\text{OD} \ll |\mathcal{P}| Q^{2+\varepsilon} (X^{1/2}/Q)^{1/2} = Q^{3/2+\varepsilon} X^{1/4} |\mathcal{P}|.$$

In either case use  $Q \leq X^{1/2-\kappa}$  and  $|\mathcal{P}| \asymp P/\log P = X^{\vartheta+o(1)}$  to obtain

$$\text{OD} \ll X^{1-\delta+\varepsilon} |\mathcal{P}|^{2-\delta} \quad \text{with} \quad \delta \leq \min \left\{ 2\kappa + \frac{\vartheta}{2}, \vartheta + \frac{3}{2}\kappa \right\}.$$

Fix

$$\delta := \frac{1}{1000} \min \left\{ \kappa, \frac{1}{2} - 3\vartheta \right\},$$

which is positive provided  $\vartheta < \frac{1}{6} - \kappa$ . Since  $Q^2 \leq X$ , we may rewrite  $X^{1-\delta} \asymp (Q^2 + X)^{1-\delta}$ , giving the stated  $\text{OD} \ll (Q^2 + X)^{1-\delta} |\mathcal{P}|^{2-\delta} X^{\varepsilon}$ .

*Step 5: Divide out the amplifier.* By the amplifier domination at the start,

$$\sum_{q, \chi, f} |S_{q, \chi, f}|^2 \leq \frac{1}{|\mathcal{P}|^2} \text{OD} \ll (Q^2 + X)^{1-\delta} |\mathcal{P}|^{-\delta} X^{\varepsilon}.$$

Taking any fixed  $\vartheta > 0$  allowed above makes  $|\mathcal{P}| = X^{\vartheta+o(1)}$ , and we absorb  $|\mathcal{P}|^{-\delta}$  into  $X^{\varepsilon}$  by shrinking  $\varepsilon$ . This yields

$$\sum_{q \in \mathcal{Q}} \sum_{\chi} \sum_f \left| \sum_{n \asymp X} \alpha_n \lambda_f(n) \chi(n) \right|^2 \ll (Q^2 + X)^{1-\delta} X^{\varepsilon},$$

uniformly across all spectral pieces, completing the proof.  $\square$

*Remark C.9* (Parameters & ranges for Lemma C.8). Fix any  $0 < \kappa < \frac{1}{4}$  and choose  $\vartheta$  with

$$0 < \vartheta < \frac{1}{6} - \kappa.$$

Take  $Q \leq X^{1/2-\kappa}$  and  $P = X^{\vartheta}$  (so  $|\mathcal{P}| \asymp P/\log P$ ). Then Lemma C.8 holds with

$$\delta = \frac{1}{1000} \min \left\{ \kappa, \frac{1}{2} - 3\vartheta \right\} > 0.$$

In particular, the choice

$$\kappa = 10^{-3}, \quad \vartheta = \frac{\kappa}{8}$$

gives  $\delta \geq 5 \times 10^{-7}$ , which is uniform across all dyadic  $X$  and all spectral pieces (holomorphic, Maaß, and Eisenstein, including oldforms). The constants in the bound

$$\sum_{q \in \mathcal{Q}} \sum_{\chi} \sum_f \left| \sum_{n \asymp X} \alpha_n \lambda_f(n) \chi(n) \right|^2 \ll (Q^2 + X)^{1-\delta} X^{\varepsilon}$$

depend at most on  $\varepsilon$ , on finitely many derivatives of the fixed test  $h$ , and on the exponent  $C$  in the divisor-type bound  $|\alpha_n| \ll_{\varepsilon} \tau(n)^C$ .

## C.2 Type-III Spectral Bound

Let  $(\alpha_n)$  be a smooth Type-III coefficient sequence supported on  $n \asymp X$ , with divisor-type bounds  $|\alpha_n| \ll_\varepsilon \tau(n)^C$  and smooth weight of width  $X^{1+o(1)}$ . For  $Q \geq 1$ , let the outer sums range over moduli  $q \leq Q$ , primitive characters  $\chi \pmod{q}$ , and an orthonormal Hecke basis  $f$  (holomorphic + Maaß, including oldforms and Eisenstein as in Kuznetsov). Assume **PASSG (Prime-averaged short-shift gain)** holds with some fixed  $\delta > 0$ . Then, for any  $\varepsilon > 0$ ,

**Proposition C.10** (Type-III spectral second moment). *Let  $(\alpha_n)$  be a smooth Type-III coefficient sequence supported on  $n \asymp X$ , with divisor-type bounds  $|\alpha_n| \ll_\varepsilon \tau(n)^C$  and smooth weight of width  $X^{1+o(1)}$ . For  $Q \geq 1$ , let the outer sums range over moduli  $q \leq Q$ , primitive characters  $\chi \pmod{q}$ , and an orthonormal Hecke basis  $f$  (holomorphic + Maaß, including oldforms and Eisenstein as in Kuznetsov). Assume **PASSG (Prime-averaged short-shift gain)** holds with some fixed  $\delta > 0$ . Then, for any  $\varepsilon > 0$ ,*

$$\sum_{q \leq Q} \sum_{\chi \pmod{q}} \sum_f \left| \sum_{n \asymp X} \alpha_n \lambda_f(n) \chi(n) \right|^2 \ll_{\varepsilon, C} (Q^2 + X)^{1-\delta} X^\varepsilon.$$

*Proof using Lemma C.8. Step 1: Balanced prime amplifier that kills the diagonal.* Let  $\mathcal{P}$  be the set of primes  $p \in [P, 2P]$  with  $P = X^\vartheta$  (to be chosen; PASSG is uniform in  $P$ ). Choose deterministic signs  $\varepsilon_p \in \{\pm 1\}$  so that

$$\sum_{p \in \mathcal{P}} \varepsilon_p = 0 \quad \text{and} \quad \left| \sum_{p \in \mathcal{P}} \varepsilon_p \varepsilon_{p+\Delta} \right| \ll |\mathcal{P}| \cdot \mathbf{1}_{|\Delta| \leq P^{1-o(1)}},$$

i.e. a “balanced Rademacher” choice; a random choice satisfies this with probability  $\gg 1$ , and we fix one such choice.

Define the amplifier on the spectrum:

$$A_f := \sum_{p \in \mathcal{P}} \varepsilon_p \lambda_f(p).$$

Because  $\sum_p \varepsilon_p = 0$ , expanding  $|A_f|^2$  removes the pure diagonal  $p = p'$  on average over signs, leaving only short prime shifts  $p \neq p'$  with  $\Delta = p - p'$  (the “short-shift” structure needed for PASSG).

**Step 2: Diagonal-free reduction by polarization.** For any complex numbers  $S_f$ ,

$$\sum_f |S_f|^2 = \frac{1}{\sum_{p \in \mathcal{P}} \varepsilon_p^2} \sum_f |S_f|^2 \cdot \left( \sum_{p \in \mathcal{P}} \varepsilon_p^2 \right) = \frac{1}{|\mathcal{P}|} \sum_f |S_f|^2 \cdot \sum_{p \in \mathcal{P}} 1.$$

Insert  $1 = \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} \varepsilon_p^2$  and then *complete the square* with  $A_f$ :

$$\sum_f |S_f|^2 = \frac{1}{|\mathcal{P}|^2} \sum_f |S_f|^2 \cdot \sum_{p, p' \in \mathcal{P}} \varepsilon_p \varepsilon_{p'} \lambda_f(p) \lambda_f(p') \leq \frac{1}{|\mathcal{P}|^2} \sum_f |A_f S_f|^2,$$

where the inequality is Cauchy-Schwarz in  $\sum_{p, p'}$  (this is the standard “balanced-amplifier domination”: the diagonal  $p = p'$  having zero mean is what prevents a trivial loss).

Apply this with

$$S_{q, \chi, f} := \sum_{n \asymp X} \alpha_n \lambda_f(n) \chi(n).$$

Summing over  $q \leq Q, \chi$  gives

$$\sum_{q \leq Q} \sum_{\chi} \sum_f |S_{q, \chi, f}|^2 \leq \frac{1}{|\mathcal{P}|^2} \sum_{q \leq Q} \sum_{\chi} \sum_f |A_f S_{q, \chi, f}|^2. \quad (\text{C.5})$$

**Step 3: Kuznetsov after opening the amplifier.** Open  $|A_f S_{q, \chi, f}|^2$  and use Hecke relations to rewrite prime factors  $\lambda_f(p) \lambda_f(n)$  as a (short) combination of  $\lambda_f(pn)$  and  $\lambda_f(n/p)$  (the latter is

discarded as  $p \nmid n$  for Type-III supports). After summing over  $(q, \chi, f)$  and applying Kuznetsov (including oldforms + Eisenstein), the contribution splits into:

- **Short-shift off-diagonal (OD):** correlations of the form  $\sum_{p \neq p' \in \mathcal{P}} \varepsilon_p \varepsilon_{p'} \sum_{m, n \asymp X} \alpha_m \overline{\alpha_n} \mathcal{K}_q(m, n; p - p')$ , with Kloosterman sums  $S(m, n; cq)$  and Bessel kernels;
- **(Spectral) diagonal/main terms:** the parts that would arise from  $p = p'$  or  $\Delta = 0$ , but these are annihilated by  $\sum_p \varepsilon_p = 0$  and by our balanced-sign choice, leaving at most lower-order boundary terms absorbed in  $X^\varepsilon$ .

Precisely this OD piece is what **PASSG** estimates *after* the amplifier and Kuznetsov:

**PASSG (assumed).** Uniformly in  $P = X^\vartheta$ ,

$$\text{OD} \ll (Q^2 + X)^{1-\delta} |\mathcal{P}|^{2-\delta} X^\varepsilon.$$

All Bessel-kernel ranges (small/large) are handled there; Weil bounds for  $S(\cdot, \cdot; \cdot)$ , the  $c \equiv 0 \pmod{q}$  constraint, oldforms and Eisenstein, and the short-shift averaging in  $\Delta$  are already accounted for in the statement of S2.4.

Therefore,

$$\sum_{q \leq Q} \sum_{\chi} \sum_f |A_f S_{q, \chi, f}|^2 \ll (Q^2 + X)^{1-\delta} |\mathcal{P}|^{2-\delta} X^\varepsilon. \quad (\text{C.6})$$

**Step 4: Divide out the amplifier and optimize  $P$ .** Insert (C.6) into (C.5):

$$\sum_{q \leq Q} \sum_{\chi} \sum_f |S_{q, \chi, f}|^2 \ll \frac{1}{|\mathcal{P}|^2} (Q^2 + X)^{1-\delta} |\mathcal{P}|^{2-\delta} X^\varepsilon = (Q^2 + X)^{1-\delta} |\mathcal{P}|^{-\delta} X^\varepsilon.$$

Choose any fixed  $\vartheta > 0$  (e.g.  $\vartheta = \delta/4$ ) so that  $|\mathcal{P}| = P/\log P = X^{\vartheta+o(1)}$  and absorb  $|\mathcal{P}|^{-\delta} = X^{-\vartheta\delta+o(1)}$  into  $X^\varepsilon$  (by shrinking  $\varepsilon$ ). This yields

$$\sum_{q \leq Q} \sum_{\chi} \sum_f \left| \sum_{n \asymp X} \alpha_n \lambda_f(n) \chi(n) \right|^2 \ll (Q^2 + X)^{1-\delta} X^\varepsilon,$$

as claimed.  $\square$

$\square$

### C.2.1 Remarks

- **Uniformity & hypotheses.** The argument only used (i) Type-III structure (smooth  $\alpha_n$ , divisor bounds), (ii) balanced prime amplifier with  $\sum \varepsilon_p = 0$ , (iii) Kuznetsov with full continuous and oldform ranges, and (iv) PASSG's OD estimate. No further spectral gap input is needed beyond what S2.4 encapsulates.
- **Why the diagonal doesn't spoil the saving.** The balanced amplifier removes the dangerous  $p = p'$  contribution *before* applying Kuznetsov. What remains are genuinely shifted correlations ( $\Delta \neq 0$ ), to which S2.4 applies and gives the  $(Q^2 + X)^{1-\delta}$  saving.
- **Choice of  $\vartheta$ .** Any fixed  $\vartheta \in (0, 1/2)$  permitted by S2.4 works; the  $|\mathcal{P}|^{-\delta}$  factor improves the exponent, and we simply absorb it into  $X^\varepsilon$ .

This completes Part C.5 once PASSG is rigorously in place.

## Part D

# Assembly

## D.1 Dyadic Decomposition (final)

### D.1.1 Statement

Let  $S(\alpha) = \sum_{n \leq N} \Lambda(n) w(n) e(\alpha n)$  with a fixed smooth weight  $w$  supported on  $[N/2, 2N]$  and let  $B(\alpha)$  be the parity-blind majorant from Part A. For the minor arcs  $\mathfrak{m}$  defined with denominator cutoff  $Q = N^{1/2-\varepsilon}$ , assume the analytic inputs:

- **(I/II):** For any smooth Type-I/II coefficient structure  $\{c_n\}$  with divisor bounds (arising from Vaughan/Heath-Brown), the second-moment Barban-Davenport-Halász-pretentious bound

$$\sum_{q \leq Q} \sum_{\chi \bmod q} \left| \sum_{n \leq N} c_n \lambda(n) \chi(n) \right|^2 \ll \frac{NQ}{(\log N)^A} \quad (\text{D.1})$$

holds for each fixed  $A > 0$ . (This is BVP2M and the “Route B Lemma” for the balanced ranges.)

- **(III):** For every dyadic Type-III block  $\sum_{n \asymp X} \alpha_n \lambda_f(n) \chi(n)$  produced after amplification and Kuznetsov, the prime-averaged off-diagonal is bounded by

$$\text{OD} \ll (Q^2 + X)^{1-\delta} |\mathcal{P}|^{2-\delta} \quad (\text{D.2})$$

for some fixed  $\delta > 0$ , uniformly for amplifier length  $|\mathcal{P}| = X^\vartheta$  with  $\vartheta = \vartheta(\delta) > 0$ , and with uniform control of oldforms/Eisenstein and Bessel kernels. (This is PASSG and its Type-III spectral corollary.)

Then, for any  $\varepsilon > 0$ ,

$$\int_{\mathfrak{m}} |S(\alpha) - B(\alpha)|^2 d\alpha \ll \frac{N}{(\log N)^{3+\varepsilon}}.$$

### D.1.2 Proof

**Step 1: Identity and dyadic model.** Apply a 3-, 4-, or 5-fold Heath-Brown identity (any standard version suffices) to  $\Lambda$  with cut parameters

$$U = N^\mu, \quad V = N^\nu, \quad W = N^\omega, \quad 0 < \mu \leq \nu \leq \omega < 1,$$

chosen below. We write

$$S(\alpha) - B(\alpha) = \sum_{\text{HB terms } \mathcal{T}} \mathcal{S}_{\mathcal{T}}(\alpha),$$

where each  $\mathcal{S}_{\mathcal{T}}$  is a finite linear combination (with coefficients having  $\ll_\varepsilon n^\varepsilon$  divisor bounds and smooth dyadic cutoffs) of exponential sums of one of the three structural types:

- **Type I:**  $\sum_{m \asymp M} a_m \sum_{n \asymp N/M} b_n e(\alpha mn)$  with  $M \leq U$  (or the dual small variable),
- **Type II:** balanced  $\sum_{m \asymp M} \sum_{n \asymp N/M} a_m b_n e(\alpha mn)$  with  $U \ll M \ll N/U$ ,
- **Type III:** “ternary” or highly factorized pieces with all variables in ranges  $\ll N^{1/3+o(1)}$ , which, after the amplifier/Kuznetsov transition, become prime-averaged short-shift sums against automorphic coefficients.

All sums are partitioned into  $\mathbf{O}((\log N)^C)$  dyadic blocks in all active variables for some fixed  $C$ .

**Step 2: Minor-arc  $L^2$  via large sieve on dyadics.** Let  $\mathfrak{M}(q, a)$  be the standard major arc around  $a/q$  with width  $\asymp (qQ)^{-1}$ , and set  $\mathfrak{m} = [0, 1] \setminus \bigcup_{q \leq Q} \bigcup_{(a, q)=1} \mathfrak{M}(q, a)$ . On  $\mathfrak{m}$  we use the standard large-sieve/dispersion reduction:

for suitable coefficients  $c_n$  associated to the dyadic block  $\mathcal{T}$ . By opening the square and expanding in Dirichlet characters modulo  $q$ , (D.2) reduces to sums of the form

$$\sum_{q \leq Q} \sum_{\chi \bmod q} \left| \sum_{n \asymp X} c_n \lambda(n) \chi(n) \right|^2 \quad (\text{D.3})$$

or, in the Type-III case after the amplifier/Kuznetsov step, to a spectral second moment whose diagonal/off-diagonal split is controlled by (D.2).

We now bound (D.3) block-wise and then sum the dyadics.

### D.1.3 Step 3: Type I/II dyadics

Choose  $U = N^{1/3}$  (any  $\mu \in (1/4, 1/2)$  is fine) so that all Type I/II ranges from the chosen Heath-Brown identity fall either in the “small-large” or “balanced” regimes. By the input (I/II), for any  $A > 0$ ,

$$\sum_{q \leq Q} \sum_{\chi \bmod q} \left| \sum_{n \leq N} c_n \lambda(n) \chi(n) \right|^2 \ll \frac{NQ}{(\log N)^A}.$$

Each Type I or Type II dyadic contributes  $\ll NQ/(\log N)^A$ . There are  $\ll (\log N)^C$  such dyadics in total, so by taking  $A \geq 3 + C + 10\varepsilon^{-1}$  we obtain

$$\sum_{\text{Type I/II dyadics}} \int_{\mathfrak{m}} |\mathcal{S}_{\mathcal{T}}(\alpha)|^2 d\alpha \ll \frac{N}{(\log N)^{3+\varepsilon}}. \quad (\text{D.4})$$

### D.1.4 Step 4: Type III dyadics

Fix  $V = W = N^{1/3}$  so that the residual blocks with all variables  $\ll N^{1/3+o(1)}$  are designated Type III. For such a block, let its “outer scale” be  $X \asymp N^\xi$  with  $\xi \in (0, 1)$  determined by the product of the active variables. After applying the amplifier of length  $|\mathcal{P}| = X^\vartheta$  and Kuznetsov, we face a spectral second moment whose off-diagonal obeys (D.2):

$$\text{OD} \ll (Q^2 + X)^{1-\delta} |\mathcal{P}|^{2-\delta} = (Q^2 + X)^{1-\delta} X^{\vartheta(2-\delta)}.$$

Take  $\vartheta = \frac{\delta}{8}$  (any fixed small choice depending on  $\delta$  works). Since  $Q = N^{1/2-\varepsilon}$ , we have  $Q^2 = N^{1-2\varepsilon}$ . Two regimes:

- If  $X \leq Q^2$  then  $\text{OD} \ll N^{(1-2\varepsilon)(1-\delta)} X^{\vartheta(2-\delta)}$ .
- If  $X \geq Q^2$  then  $\text{OD} \ll X^{1-\delta+\vartheta(2-\delta)}$ .

In both cases there is a fixed saving  $X^{-\eta}$  (or  $N^{-\eta}$ ) for some  $\eta = \eta(\delta, \vartheta, \varepsilon) > 0$  against the trivial diagonal scale, after the standard dispersion normalization. Consequently each Type III dyadic contributes

$$\int_{\mathfrak{m}} |\mathcal{S}_{\mathcal{T}}(\alpha)|^2 d\alpha \ll \frac{N}{(\log N)^A} X^{-\eta} + (\text{diagonal}). \quad (\text{D.5})$$

The diagonal is controlled either by the amplifier normalization or by subtracting the parity-blind majorant  $B(\alpha)$  (which removes the main term on  $\mathfrak{m}$ ), leaving at most  $\ll N/(\log N)^A$  per block. Summing (D.5) over the  $\ll (\log N)^C$  Type-III dyadics and choosing  $A$  large, we obtain

$$\sum_{\text{Type III dyadics}} \int_{\mathfrak{m}} |\mathcal{S}_{\mathcal{T}}(\alpha)|^2 d\alpha \ll \frac{N}{(\log N)^{3+\varepsilon}}. \quad (\text{D.6})$$



*Bookkeeping note.* The  $X^{-\eta}$  saving is uniform in the dyadic location because  $\delta > 0$  is fixed and  $\vartheta$  is chosen as a fixed fraction of  $\delta$ ; any residual factors from Bessel kernels, oldforms, and Eisenstein are already absorbed in (D.2) by the uniform spectral analysis ensured in PASSG. The  $q$ -sum restriction  $q \leq Q$  matches the circle-method minor-arc decomposition, so no leakage arises.

### D.1.5 Step 5: Conclusion

Adding (D.4) and (D.6) over all dyadics of all HB terms  $\mathcal{T}$  yields

$$\int_{\mathfrak{m}} |S(\alpha) - B(\alpha)|^2 d\alpha \ll \frac{N}{(\log N)^{3+\varepsilon}},$$

as claimed.

### D.1.6 Derivation of (A.1) from BVP2M and PASSGs

**Scope.** In this subsection we *derive* the minor-arc  $L^2$  estimate

$$\int_{\mathfrak{m}} |S(\alpha) - B(\alpha)|^2 d\alpha \ll \frac{N}{(\log N)^{3+\varepsilon}}$$

(i) **Type I/II second moment with parity** (BVP2M): for  $Q \leq N^{1/2}(\log N)^{-B}$ ,

$$\sum_{q \leq Q} \sum_{\chi \bmod q} \left| \sum c_n \lambda(n) \chi(n) \right|^2 \ll \frac{NQ}{(\log N)^A},$$

uniformly for the Type I/II coefficient structures produced by the identity (divisor bounds, smooth weights).

(ii) **Type III off-diagonal saving** (PASSG): after prime-length amplification and Kuznetsov,

$$\text{OD} \ll (Q^2 + X)^{1-\delta} |\mathcal{P}|^{2-\delta} X^\varepsilon$$

for some fixed  $\delta > 0$  (with  $|\mathcal{P}| = X^\vartheta$ ,  $0 < \vartheta < \frac{1}{6} - \kappa$ ), uniformly across spectral families.

**Large-sieve reduction on  $\mathfrak{m}$ .** For each Heath-Brown dyadic block  $\mathcal{T}$ , Gallagher's/large-sieve minor-arc reduction (Lemma D.1) yields

$$\int_{\mathfrak{m}} |\mathcal{S}_{\mathcal{T}}(\alpha)|^2 d\alpha \ll \frac{1}{Q^2} \sum_{q \leq Q} \sum_{\substack{a \bmod q \\ (a,q)=1}} \left| \sum_n c_n e\left(\frac{an}{q}\right) \right|^2.$$

Expanding in Dirichlet characters reduces this to the second moments controlled by (i) and (ii).

**Type I/II dyadics.** BVP2M with  $A$  large (absorbing the  $O((\log N)^C)$  dyadic inflation) gives a total

$$\sum_{\text{Type I/II dyadics}} \int_{\mathfrak{m}} |\mathcal{S}_{\mathcal{T}}(\alpha)|^2 d\alpha \ll \frac{N}{(\log N)^{3+\varepsilon}}.$$

**Type III dyadics.** After applying the prime amplifier of fixed length  $|\mathcal{P}| = X^\vartheta$  and Kuznetsov, PASSG furnishes a uniform saving  $\delta > 0$  on the off-diagonal. Dividing by the amplifier normalization (as in Prop. C.10), one gets for each Type III block (with outer scale  $X$ )

$$\int_{\mathfrak{m}} |\mathcal{S}_{\mathcal{T}}(\alpha)|^2 d\alpha \ll Q^{-2} (Q^2 + X)^{1-\delta} X^{-\vartheta\delta+\varepsilon}.$$

Summing over Type III dyadics and splitting  $X \leq Q^2$  and  $X \geq Q^2$  yields a net contribution  $\ll N(\log N)^{-3-\varepsilon}$  for fixed  $\vartheta = \vartheta(\delta) > 0$ .

**Conclusion.** Summing all dyadics gives (A.1). *Thus, (A.1) holds provided BVP2M and PASSG hold in the stated uniform forms.* This is the only place where (A.1) depends on Part B and Part C.

### D.1.7 Parameter choices & loss ledger (for ease of cross-checking)

- **Minor-arc cutoff:**  $Q = N^{1/2-\varepsilon}$ .
- **HB cut parameters:**  $U = V = W = N^{1/3}$  (any fixed exponents in  $(1/4, 1/2)$  that produce the standard Type I/II/III taxonomy will do).
- **Amplifier:** primes of length  $|\mathcal{P}| = X^\vartheta$  with  $\vartheta = \delta/8$ .
- **Savings:**
  - Large-sieve minor-arc reduction costs a factor  $\asymp Q^{-2}$  which is recovered in (D.1)/(D.2).
  - Type I/II: pick  $A$  so that  $(\log N)^C$  dyadic inflation is dominated; we target  $3+\varepsilon$  net powers of log.
  - Type III: the  $\delta$ -saving from (D.2) after amplifier normalization yields uniform  $X^{-\eta}$  decay, summable across dyadics.
- **Exceptional characters / oldforms / Eisenstein:** already handled in the hypotheses of BVP2M and PASSG; their contributions obey the same  $(\log N)^{-A}$  savings and therefore do not affect the sum.

### D.1.8 Remark

Nothing delicate hinges on the exact form of the identity (Vaughan vs. Heath-Brown) provided it yields (i) divisor-bounded smooth coefficients and (ii) a genuine three-variable “Type III” regime where PASSG applies. Alternative cut choices merely reshuffle a finite number of dyadic families and do not change the final  $(\log N)^{-3-\varepsilon}$  power once  $A$  is taken large in the Type I/II inputs.

## D.2 Major-Arc Evaluation

Let

$$\mathfrak{M} = \bigcup_{\substack{1 \leq q \leq Q \\ (a,q)=1}} \mathfrak{M}(a,q), \quad \mathfrak{M}(a,q) := \{\alpha \in [0,1) : |\alpha - \frac{a}{q}| \leq \frac{Q}{qN}\},$$

with  $Q = N^{1/2-\varepsilon}$ . Write  $\alpha = a/q + \beta$  on  $\mathfrak{M}(a,q)$  and set

$$V(\beta) := \sum_{n \leq N} e(n\beta) \quad \text{and} \quad \widehat{w}(\beta) := \sum_n w(n)e(n\beta)$$

for the sharp/smoothed Dirichlet kernels according to whether  $S, B$  are unweighted or carry a fixed smooth weight  $w$  supported on  $[1, N]$  with  $w^{(j)} \ll_j N^{-j}$ .

We denote by  $\mathfrak{S}(N)$  the (Goldbach) singular series

$$\mathfrak{S}(N) = 2 \prod_{p \geq 3} \left(1 - \frac{1}{(p-1)^2}\right) \prod_{\substack{p|N \\ p \geq 3}} \frac{p-1}{p-2},$$

and by  $\mathfrak{J}$  the singular integral

$$\mathfrak{J} = \begin{cases} \int_{-\infty}^{\infty} \left| \frac{\sin(\pi N \beta)}{\sin(\pi \beta)} \right|^2 e(-N\beta) d\beta & \text{(sharp cut-off),} \\ \int_{-\infty}^{\infty} |\widehat{w}(\beta)|^2 e(-N\beta) d\beta & \text{(smooth cut-off).} \end{cases}$$

Standard analysis yields  $\mathfrak{J} = N + O(1)$  in the sharp case and  $\mathfrak{J} = \widehat{w}(0)^2 N + O(1)$  in the smooth case.

We evaluate first the parity-blind majorant  $B$ , then transfer the main term to  $S$ .

### D.2.1 Major-arc evaluation for $B(\alpha)$

Let the sieve majorant be

$$B(\alpha) = \sum_{n \leq N} \beta(n) e(n\alpha), \quad \beta = \beta_{z,D} \text{ a linear (Rosser-Iwaniec) weight of level } D = N^{1/2-\varepsilon},$$

so that  $\beta$  has the standard divisor-bounded structure

$$\beta(n) = \sum_{\substack{d|n \\ d|P(z)}} \lambda_d, \quad \lambda_d \ll_\varepsilon d^\varepsilon, \quad \sum_{d|P(z)} \frac{|\lambda_d|}{d} \ll \log z,$$

with  $P(z) = \prod_{p < z} p$  and  $z = N^\eta$  a small fixed power.

On  $\alpha = a/q + \beta$  with  $q \leq Q$  and  $|\beta| \leq Q/(qN)$ , expand

$$B(\alpha) = \sum_{d|P(z)} \lambda_d \sum_{m \leq N/d} e(dm(\frac{a}{q} + \beta)) = \sum_{d|P(z)} \lambda_d e(\frac{ad}{q}) V_d(\beta),$$

where  $V_d(\beta) := \sum_{m \leq N/d} e(dm\beta)$ . By the standard completion and the Euler product calculation for linear sieve weights (matching local factors for  $p < z$ ), one obtains the **major-arc approximation**

$$B(a/q + \beta) = \frac{\rho(q)}{\varphi(q)} V(\beta) + \mathcal{E}_B(q, \beta),$$

where  $\rho(q)$  is multiplicative, supported on square-free  $q$ , and satisfies

$$\rho(p) = \begin{cases} -1 & \text{for } p \geq 3, \\ 0 & \text{for } p = 2, \end{cases} \quad \text{so that} \quad \frac{\rho(q)}{\varphi(q)} = \frac{\mu(q)}{\varphi(q)}$$

for all odd  $q$  with  $p < z$  local factors correctly matched. Moreover, uniformly for  $q \leq Q$  and  $|\beta| \leq Q/(qN)$ ,

$$\mathcal{E}_B(q, \beta) \ll N(\log N)^{-A}$$

for any fixed  $A > 0$  once  $z = N^\eta$  and  $D = N^{1/2-\varepsilon}$  are tied as usual (this is the standard “well-factorable” savings of the linear sieve on major arcs).

Squaring and integrating over  $\mathfrak{M}$  (disjoint up to negligible overlaps) gives

$$\int_{\mathfrak{M}} B(\alpha)^2 e(-N\alpha) d\alpha = \sum_{q \leq Q} \sum_{\substack{a \bmod q \\ (a,q)=1}} \int_{|\beta| \leq Q/(qN)} \left( \frac{\mu(q)}{\varphi(q)} V(\beta) \right)^2 e(-N\beta) d\beta + O\left( \frac{N}{(\log N)^{3+\varepsilon}} \right),$$

where the error uses Cauchy-Schwarz with  $\int_{\mathfrak{M}} |V(\beta)|^2 d\beta \ll N \log N$ , the uniform bound on  $\mathcal{E}_B$ , and the total measure of  $\mathfrak{M}$ . Since  $\sum_{(a,q)=1} 1 = \varphi(q)$  and  $\int_{|\beta| \leq Q/(qN)} V(\beta)^2 e(-N\beta) d\beta = \mathfrak{J} + O(NQ^{-1})$ ,

$$\int_{\mathfrak{M}} B(\alpha)^2 e(-N\alpha) d\alpha = \left( \sum_{q=1}^{\infty} \frac{\mu(q)^2}{\varphi(q)^2} c_q(N) \right) \mathfrak{J} + O\left( \frac{N}{(\log N)^{3+\varepsilon}} \right),$$

with  $c_q(N)$  the Ramanujan sum. The absolutely convergent series equals the Goldbach singular series  $\mathfrak{S}(N)$ . Hence

$$\boxed{\int_{\mathfrak{M}} B(\alpha)^2 e(-N\alpha) d\alpha = \mathfrak{S}(N) \mathfrak{J} + O(N(\log N)^{-3-\varepsilon}) .}$$

*Remark.* If a smooth weight  $w$  is used, replace  $V(\beta)$  by  $\widehat{w}(\beta)$  throughout, and the same argument yields  $\mathfrak{J} = \int |\widehat{w}|^2 e(-N\beta) d\beta$  with an identical error term.

### D.2.2 Transferring the main term to $S(\alpha)$

Let  $S(\alpha) = \sum_{n \leq N} \Lambda(n) e(n\alpha)$  (sharp or smooth as above). By the prime number theorem in arithmetic progressions with level of distribution  $Q = N^{1/2-\varepsilon}$  (Siegel-Walfisz + Bombieri-Vinogradov in the smooth form used earlier), uniformly for  $q \leq Q$  and  $|\beta| \leq Q/(qN)$ ,

$$S(a/q + \beta) = \frac{\mu(q)}{\varphi(q)} V(\beta) + \mathcal{E}_S(q, \beta), \quad \mathcal{E}_S(q, \beta) \ll N(\log N)^{-A}$$

for any fixed  $A > 0$ . Consequently, exactly the same computation as in §7.1 gives

$$\int_{\mathfrak{M}} S(\alpha)^2 e(-N\alpha) d\alpha = \mathfrak{S}(N) \mathfrak{J} + O(N(\log N)^{-3-\varepsilon}).$$

There are two convenient “comparison” routes:

- **Pointwise on  $\mathfrak{M}$ :** From the two approximations above,

$$S(\alpha) - B(\alpha) = \mathcal{E}_S(\alpha) - \mathcal{E}_B(\alpha),$$

whence  $\int_{\mathfrak{M}} (S^2 - B^2) e(-N\alpha) d\alpha = \int_{\mathfrak{M}} (S - B)(S + B) e(-N\alpha) d\alpha$  is  $\ll N(\log N)^{-A}$  after the same bookkeeping.

- **Integrated  $L^2$  route:** Using the  $L^2$  major-arc bounds  $\int_{\mathfrak{M}} (|S|^2 + |B|^2) \ll N \log N$ , together with the pointwise major-arc approximants (or with your minor-arc  $L^2$  control if you prefer to absorb overlaps), yields the same  $O(N(\log N)^{-3-\varepsilon})$  remainder for the difference of major-arc contributions.

Combining §7.1-§7.2 we conclude the following proposition.

**Proposition 7.1 (Major-arc main term).** For the major arcs  $\mathfrak{M}$  with  $Q = N^{1/2-\varepsilon}$ ,

$$\int_{\mathfrak{M}} B(\alpha)^2 e(-N\alpha) d\alpha = \int_{\mathfrak{M}} S(\alpha)^2 e(-N\alpha) d\alpha = \mathfrak{S}(N) \mathfrak{J} + O(N(\log N)^{-3-\varepsilon}).$$

In particular,  $B$  and  $S$  share the same Hardy-Littlewood main term on the major arcs, with an error that is negligible against  $N(\log N)^{-2}$ .

### Completion of the Minor-Arc Analysis

#### Derivation of (A.1) from Lemma B.2 and Lemma C.8

We now give a compact, self-contained deduction of the minor-arc bound

$$\boxed{\int_{\mathfrak{m}} |S(\alpha) - B(\alpha)|^2 d\alpha \ll \frac{N}{(\log N)^{3+\varepsilon}}}, \quad (\text{A.1})$$

using only Lemma B.2 (Type I/II second moment with parity) and Lemma C.8 (prime-averaged short-shift gain for Type III).

**Setup and parameters.** Fix  $\varepsilon \in (0, 10^{-2})$  and set  $Q = N^{1/2-\varepsilon}$  for the major/minor arc decomposition. Apply a Heath-Brown identity with symmetric cuts  $U = V = W = N^{1/3}$  to  $\Lambda$  in  $S(\alpha)$ , and subtract the parity-blind majorant  $B(\alpha)$  (linear/Rosser-Iwaniec sieve at level  $D = N^{1/2-\varepsilon}$ ). This yields

$$S(\alpha) - B(\alpha) = \sum_{\mathcal{T}} \mathcal{S}_{\mathcal{T}}(\alpha),$$

where the finitely many  $\mathcal{T}$  are dyadic Type I/II/III blocks with divisor-bounded smooth coefficients supported on  $n \asymp X$  for some  $X$ .

**Minor-arc large-sieve reduction.** For each block  $\mathcal{T}$  with coefficient sequence  $c_n$  (carrying the smooth dyadics), Gallagher's minor-arc reduction (Lemma D.1) gives

$$\int_{\mathfrak{m}} \left| \sum_n c_n e(\alpha n) \right|^2 d\alpha \ll \frac{1}{Q^2} \sum_{q \leq Q} \sum_{\substack{a \pmod{q} \\ (a,q)=1}} \left| \sum_n c_n e\left(\frac{an}{q}\right) \right|^2.$$

Expanding in Dirichlet characters mod  $q$  reduces this to second moments of the shape

$$\sum_{q \leq Q} \sum_{\chi \pmod{q}} \left| \sum_{n \asymp X} c_n \chi(n) \right|^2,$$

with the *parity twist*  $\lambda(n)$  present inside  $c_n$  for the terms arising from  $S - B$ .

**Type I/II blocks.** By Lemma B.2 (with  $Q \leq N^{1/2}(\log N)^{-B}$  and  $L \geq N^\eta$  whenever needed),

$$\sum_{q \leq Q} \sum_{\chi \pmod{q}} \left| \sum_{n \asymp X} c_n \lambda(n) \chi(n) \right|^2 \ll \frac{XQ}{(\log N)^A}.$$

Summed over the  $O((\log N)^C)$  Type I/II dyadics (with  $X \asymp N$  up to constants), and multiplied by the prefactor  $Q^{-2}$  from the minor-arc reduction, this yields

$$\sum_{\text{Type I/II}} \int_{\mathfrak{m}} |\mathcal{S}_{\mathcal{T}}(\alpha)|^2 d\alpha \ll \frac{N}{(\log N)^{3+\varepsilon}},$$

upon taking  $A$  large enough in terms of  $C$  and  $\varepsilon$ .

**Type III blocks.** For a Type III block at outer scale  $X$ , apply the balanced prime amplifier and Kuznetsov as in Part C to reach the spectral second moment controlled by Lemma C.8. With  $P = X^\vartheta$  (any fixed  $\vartheta$  with  $0 < \vartheta < \frac{1}{6} - \kappa$ ) and  $Q \leq X^{1/2-\kappa}$ , Lemma C.8 gives

$$\sum_{q \leq Q} \sum_{\chi} \sum_f \left| \sum_{n \asymp X} \alpha_n \lambda_f(n) \chi(n) \right|^2 \ll (Q^2 + X)^{1-\delta} X^\varepsilon, \quad \delta = \frac{1}{1000} \min\{\kappa, \tfrac{1}{2} - 3\vartheta\} > 0.$$

Dividing out the amplifier (as in Lemma C.8) and undoing the spectral expansion (orthogonality), one obtains for each Type III block

$$\sum_{q \leq Q} \sum_{\chi \pmod{q}} \left| \sum_{n \asymp X} c_n \lambda(n) \chi(n) \right|^2 \ll (Q^2 + X)^{1-\delta} X^\varepsilon.$$

Inserting this into the minor-arc large-sieve reduction yields

$$\int_{\mathfrak{m}} |\mathcal{S}_{\mathcal{T}}(\alpha)|^2 d\alpha \ll Q^{-2} (Q^2 + X)^{1-\delta} X^\varepsilon.$$

Summing over the  $O((\log N)^C)$  Type III dyadics and splitting into  $X \leq Q^2$  and  $X \geq Q^2$  gives a uniform power saving:

$$\sum_{\text{Type III}} \int_{\mathfrak{m}} |\mathcal{S}_{\mathcal{T}}(\alpha)|^2 d\alpha \ll \frac{N}{(\log N)^{3+\varepsilon}},$$

since  $(Q^2 + X)^{1-\delta} Q^{-2} \leq Q^{-2\delta}$  when  $X \leq Q^2$ , and  $\leq X^{-\delta}$  when  $X \geq Q^2$ , both summable over dyadics (choose  $\kappa, \vartheta$  once for all dyadics so that  $\delta > 0$ ).

**Conclusion.** Adding Type I/II and Type III contributions and recalling  $S - B = \sum_{\mathcal{T}} \mathcal{S}_{\mathcal{T}}$ , we obtain (A.1). All constants depend at most on  $\varepsilon$  (the minor-arc width), on the fixed smooth cutoff in the Heath–Brown identity, on  $k$  and the divisor-type bounds for coefficients, and on finitely many derivatives of the fixed Kuznetsov test  $h$ .  $\square$

### D.2.3 Status

Everything here is standard Hardy-Littlewood major-arc analysis. What remains (and is already ensured by our earlier sections) is to (i) state the exact sieve parameters  $(z, D)$  used to define  $\beta$ , and (ii) cite the precise Bombieri-Vinogradov/Siegel-Walfisz input in the smooth form employed so the uniform error  $N(\log N)^{-A}$  on  $\mathfrak{M}$  holds (both for  $\Lambda$  and for the linear-sieve majorant).

## D.3 Final Step (conditional on (A.1))

We now conclude the argument.

$$R(N) = \int_0^1 S(\alpha)^2 e(-N\alpha) d\alpha = \int_{\mathfrak{M}} S(\alpha)^2 e(-N\alpha) d\alpha + \int_{\mathfrak{m}} S(\alpha)^2 e(-N\alpha) d\alpha.$$

### D.3.1 Major arcs

By the Major-Arc Evaluation (Part D.7), we have, uniformly for even  $N$ ,

$$\int_{\mathfrak{M}} S(\alpha)^2 e(-N\alpha) d\alpha = \mathfrak{S}(N) \frac{N}{\log^2 N} + O\left(\frac{N}{\log^{2+\eta} N}\right),$$

for some fixed  $\eta > 0$ . Here  $\mathfrak{S}(N)$  is the binary Goldbach singular series

$$\mathfrak{S}(N) = 2 \prod_{p \geq 3} \left(1 - \frac{1}{(p-1)^2}\right) \prod_{\substack{p|N \\ p \geq 3}} \left(1 + \frac{1}{p-2}\right),$$

which satisfies  $\mathfrak{S}(N) > 0$  for every even  $N$ , and  $\mathfrak{S}(N) = 0$  for odd  $N$ .

### D.3.2 Minor arcs

Assume the minor-arc  $L^2$  input (A.1):

$$\int_{\mathfrak{m}} |S(\alpha) - B(\alpha)|^2 d\alpha \ll \frac{N}{(\log N)^{3+\varepsilon}}.$$

Write  $S^2 = B^2 + 2B(S - B) + (S - B)^2$  and integrate over  $\mathfrak{m}$ . By Cauchy-Schwarz and Parseval,

$$\left| \int_{\mathfrak{m}} B(\alpha) (S(\alpha) - B(\alpha)) e(-N\alpha) d\alpha \right| \leq \left( \int_0^1 |B(\alpha)|^2 d\alpha \right)^{1/2} \left( \int_{\mathfrak{m}} |S(\alpha) - B(\alpha)|^2 d\alpha \right)^{1/2} \ll \frac{N}{(\log N)^{2+\varepsilon/2}},$$

since  $\int_0^1 |B|^2 \ll N/\log N$  by (B2)-(B3). The pure error  $\int_{\mathfrak{m}} |S - B|^2$  is already  $\ll N/(\log N)^{3+\varepsilon}$ . Thus the minor arcs contribute  $o(N/\log^2 N)$  under (A.1), without requiring any bound stronger than  $\int_0^1 |B|^2 \ll N/\log N$ .

### D.3.3 Conclusion

Combining the two ranges,

$$R(N) = \mathfrak{S}(N) \frac{N}{\log^2 N} + o\left(\frac{N}{\log^2 N}\right).$$

Since  $\mathfrak{S}(N) > 0$  for every even  $N$ , it follows that  $R(N) > 0$  for all sufficiently large even  $N$ . Hence **every sufficiently large even integer is a sum of two primes.**  $\square$

### D.3.4 Remark (scope)

If desired, the error can be recorded explicitly as

$$R(N) = \mathfrak{S}(N) \frac{N}{\log^2 N} + O\left(\frac{N}{\log^{2+\eta} N}\right),$$

with the  $\eta > 0$  coming from your major-arc saving and the minor-arc  $L^2$  bound.

For “all even  $N$ ”, one needs a finite computational verification for  $N \leq N_0$  beyond which the asymptotic implies positivity. We do not specify  $N_0$  here; determining it would require explicit constants throughout (major arcs, large sieve, and spectral bounds) and numerical estimates of  $\mathfrak{S}(N)$ .

## Appendix I Technical Lemmas and Parameters

### Appendix I.1 Minor-arc large sieve reduction

We record the precise form of the inequality used in Part D.6.

**Lemma D.1** (Minor-arc large sieve reduction). *Let  $Q = N^{1/2-\varepsilon}$  and define major arcs*

$$\mathfrak{M}(q, a) = \left\{ \alpha \in [0, 1) : \left| \alpha - \frac{a}{q} \right| \leq \frac{1}{qQ} \right\}, \quad \mathfrak{M} = \bigcup_{\substack{q \leq Q \\ (a, q) = 1}} \mathfrak{M}(q, a), \quad \mathfrak{m} = [0, 1) \setminus \mathfrak{M}.$$

Then for any finitely supported sequence  $c_n$ ,

$$\int_{\mathfrak{m}} \left| \sum_n c_n e(\alpha n) \right|^2 d\alpha \ll \frac{1}{Q^2} \sum_{q \leq Q} \sum_{\substack{a \pmod{q} \\ (a, q) = 1}} \left| \sum_n c_n e\left(\frac{an}{q}\right) \right|^2.$$

*Sketch.* Partition  $[0, 1)$  into  $\{\mathfrak{M}(q, a)\}$  and  $\mathfrak{m}$ . For  $\alpha \in \mathfrak{m}$  one has  $|\alpha - \frac{a}{q}| \geq 1/(qQ)$  for all  $q \leq Q$ . Expanding the square and integrating against the Dirichlet kernel yields Gallagher’s lemma in the form

$$\int_I \left| \sum c_n e(\alpha n) \right|^2 d\alpha \ll \frac{1}{|I|^2} \sum_{q \leq 1/|I|} \sum_{a \pmod{q}} \left| \sum c_n e(an/q) \right|^2$$

for each interval  $I \subset [0, 1)$ . Applying this to each complementary arc of length  $\gg (qQ)^{-1}$  gives the stated bound.  $\square$

### Appendix I.2 Sieve weight $\beta$ and properties

Fix parameters

$$D = N^{1/2-\varepsilon}, \quad z = N^\eta \quad (0 < \eta \ll \varepsilon).$$

Let  $P(z) = \prod_{p < z} p$  and define the linear (Rosser–Iwaniec) sieve weight

$$\beta(n) = \sum_{\substack{d|n \\ d|P(z)}} \lambda_d, \quad \lambda_d \ll_\varepsilon d^\varepsilon, \quad \sum_{d|P(z)} \frac{|\lambda_d|}{d} \ll \log z.$$

**Lemma D.2.** *With this choice of  $\beta = \beta_{z, D}$  the following hold:*

(B1)  $\beta(n) \geq 0$  and  $\beta(n) \gg \frac{\log D}{\log N}$  for  $n \leq N$  almost prime.

(B2)  $\sum_{n \leq N} \beta(n) = (1 + o(1)) \frac{N}{\log N}$  and uniformly for  $(a, q) = 1$ ,  $q \leq D$ ,

$$\sum_{\substack{n \leq N \\ n \equiv a \pmod{q}}} \beta(n) = (1 + o(1)) \frac{N}{\varphi(q) \log N}.$$

(B3)  $\beta$  is well-factorable:  $\beta = \sum_{d \leq D} \lambda_d 1_{d| \cdot}$  with divisor-bounded  $\lambda_d$ , enabling major-arc analysis.

(B4) Parity-blindness. For any fixed smooth  $W$  supported on  $[1/2, 2]$ ,

$$\sum_{n \leq N} \beta(n) \lambda(n) W(n/N) \ll \frac{N}{(\log N)^A}$$

for all  $A > 0$ , uniformly in  $N$ . This follows by expanding  $\beta$ , applying Cauchy over  $d \leq D$ , and invoking BVP2M / Route B on each inner sum.

### Appendix I.3 Major-arc uniform error

**Lemma D.3** (Major-arc approximants). *Let  $\alpha = a/q + \beta$  with  $q \leq Q$ ,  $|\beta| \leq Q/(qN)$ . Then for any  $A > 0$ ,*

$$\begin{aligned} S(\alpha) &= \frac{\mu(q)}{\varphi(q)} V(\beta) + O\left(\frac{N}{(\log N)^A}\right), \\ B(\alpha) &= \frac{\mu(q)}{\varphi(q)} V(\beta) + O\left(\frac{N}{(\log N)^A}\right), \end{aligned}$$

uniformly in  $q, a, \beta$ . Here  $V(\beta) = \sum_{n \leq N} e(n\beta)$ .

*Proof.* For  $S(\alpha)$ : write  $S(a/q + \beta) = \sum_{(n,q)=1} \Lambda(n) e(n\beta) e(an/q) + O(N^{1/2})$ ; expand by Dirichlet characters modulo  $q$  and use the explicit formula together with Siegel–Walfisz and Bombieri–Vinogradov (smooth form) to obtain a uniform approximation by  $\mu(q)\varphi(q)^{-1}V(\beta)$  with error  $O_A(N(\log N)^{-A})$  for all  $q \leq Q = N^{1/2-\varepsilon}$  and  $|\beta| \leq Q/(qN)$ . See, e.g., Iwaniec–Kowalski, *Analytic Number Theory* (IK), Thm. 17.4 and Cor. 17.12, and Montgomery–Vaughan, *Multiplicative Number Theory* I.

For  $B(\alpha)$ : expand the linear (Rosser–Iwaniec) sieve weight  $\beta$  as a well-factorable convolution at level  $D = N^{1/2-\varepsilon}$ , unfold the congruences, and evaluate the major arcs via the same character expansion. The well-factorability yields savings  $O_A(N(\log N)^{-A})$  uniformly; see IK, Ch. 13 (Linear sieve; well-factorability, Thm. 13.6 and Prop. 13.10). Combining these gives the stated uniform bounds.  $\square$

### Appendix I.4 Kuznetsov at level $q$ (uniform form) and a $\Delta$ -second-moment lemma

We fix the Kuznetsov normalization we use throughout and record the uniform kernel bounds in  $q$ .

**Lemma D.4** (Kuznetsov on  $\Gamma_0(q)$  with level-uniform kernel bounds). *Let  $q \geq 1$ ,  $m, n \geq 1$  with  $(mn, q) = 1$ . For an even  $h \in C_c^\infty(\mathbb{R})$  define  $h_Q(t) := h(t/Q)$ ,  $Q \geq 1$ . Write the Kuznetsov formula on  $\Gamma_0(q)$  as*

$$\mathcal{H}_q(h_Q; m, n) = \delta_{m=n} \mathcal{D}_q(h_Q) + \sum_{c \equiv 0 \pmod{q}} \frac{1}{c} S(m, n; c) \mathcal{W}_q^{(*)} \left( \frac{4\pi\sqrt{mn}}{c} \right),$$

where  $(*) \in \{\text{Ma}\beta, \text{hol}, \text{Eis}\}$  denotes the Maaß/holomorphic/Eisenstein pieces. Then for every  $A, j \geq 0$ ,

$$\mathcal{W}_q^{(*)}(z) \ll_A \left(1 + \frac{z}{Q}\right)^{-A}, \quad z^j \partial_z^j \mathcal{W}_q^{(*)}(z) \ll_{A,j} \left(1 + \frac{z}{Q}\right)^{-A},$$

uniformly in  $q \geq 1$ ,  $z > 0$ , and in the spectral piece  $(*)$ . The implied constants depend only on  $A, j$  and  $h$  (via finitely many derivatives), not on  $q$ .

*Proof.* We record the Maaß case; the holomorphic and Eisenstein kernels are analogous. For Maaßforms the kernel is a Hankel transform

$$\mathcal{W}_q^{\text{Ma}\beta}(z) = \frac{i}{\pi} \int_{-\infty}^{\infty} h_Q(t) \tanh(\pi t) J_{2it}(z) t \, dt.$$



Since  $h \in C_c^\infty([-2, 2])$  is fixed,  $h_Q(t) = h(t/Q)$  is supported on  $|t| \leq 2Q$  and satisfies  $\|h_Q^{(r)}\|_\infty \ll_r Q^{-r}$ . Use the Schlöfli representation

$$J_{2it}(z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{iz \sin \theta} e^{-2it\theta} d\theta,$$

or, equivalently, Mellin-Barnes representations; either way, after interchanging integrals (justified by compact support) one integrates by parts in  $t$  repeatedly against the factor  $e^{-2it\theta}$ . Each  $t$ -derivative falls on  $h_Q(t) \tanh(\pi t) t$ , gaining a factor  $\ll Q^{-1}$  thanks to the  $h_Q^{(r)}$  bounds and polynomial growth control of  $\tanh$  and  $t \mapsto t$ . Thus for any  $R \geq 0$ ,

$$\mathcal{W}_q^{\text{MaB}}(z) \ll_R \int_{-\pi}^{\pi} (1 + |z \sin \theta|)^{-R} d\theta \ll_R (1 + z)^{-R}.$$

To insert the  $Q$ -scale, rescale  $t \mapsto Qt$  in the definition of  $h_Q$ ; every integration-by-parts step gains a factor  $(1 + z/Q)^{-1}$  rather than  $(1 + z)^{-1}$ , yielding  $\mathcal{W}_q^{\text{MaB}}(z) \ll_A (1 + z/Q)^{-A}$  for all  $A$ . For  $z$ -derivatives one differentiates under the integral; each  $z\partial_z$  inserts a bounded polynomial in  $t$  multiplying  $J_{2it}$  (via Bessel ODE or by differentiating the oscillatory integral), which is absorbed by the same integration-by-parts argument because  $|t| \leq 2Q$ . Uniformity in  $q$  is immediate:  $q$  appears only as the congruence condition  $c \equiv 0 \pmod{q}$  on the geometric side; it does not enter the kernel transform. The holomorphic and Eisenstein kernels are handled identically (replace  $J_{2it}$  by  $J_{k-1}$  or  $K_{2it}$ ; compact support in  $t$  gives the same decay).  $\square$

**Lemma D.5** ( $\Delta$ -second moment, level-uniform). *Let  $X \geq 3$ ,  $q \geq 1$ , and write  $c = qr$  with  $r \asymp R \geq 1$ . Fix  $P \geq 1$ . For each  $(q, r)$ , let  $W_{q,r}(m, \Delta)$  be a smooth weight supported on*

$$m \asymp X, \quad |\Delta| \leq P,$$

*with derivative bounds, for all  $0 \leq i, j \leq 10$ ,*

$$\partial_m^i \partial_\Delta^j W_{q,r}(m, \Delta) \ll_{i,j} X^{-i} P^{-j}.$$

*Define*

$$\Sigma_{q,r}(\Delta) := \sum_{m \asymp X} S(m, m + \Delta; c) W_{q,r}(m, \Delta), \quad c = qr.$$

*Then for every  $\varepsilon > 0$ ,*

$$\sum_{|\Delta| \leq P} |\Sigma_{q,r}(\Delta)|^2 \ll_\varepsilon (P + c) c^{1+2\varepsilon} X^{1+2\varepsilon},$$

*uniformly in  $q, r$  and in the family  $\{W_{q,r}\}$  subject to the stated derivative conditions.*

*Proof.* Insert a smooth dyadic cutoff  $\Psi(m/X)$  to localize  $m \in [X, 2X]$ ; absorb it into  $W_{q,r}$ . Open the square:

$$\sum_{|\Delta| \leq P} |\Sigma_{q,r}(\Delta)|^2 = \sum_{|\Delta| \leq P} \sum_{m_1, m_2 \asymp X} S(m_1, m_1 + \Delta; c) \overline{S(m_2, m_2 + \Delta; c)} W(m_1, \Delta) \overline{W(m_2, \Delta)}.$$

Expanding the Kloosterman sums gives

$$\mathcal{S} = \sum_{\substack{x_1, x_2 \bmod c \\ (x_i, c) = 1}} \sum_{|\Delta| \leq P} \sum_{m_1, m_2 \asymp X} e\left(\frac{m_1(x_1 + \bar{x}_1) - m_2(x_2 + \bar{x}_2)}{c}\right) e\left(\frac{\Delta(\bar{x}_1 - \bar{x}_2)}{c}\right) W(m_1, \Delta) \overline{W(m_2, \Delta)}.$$

*Poisson in  $\Delta$ .* Fix  $x_1, x_2$ . Writing  $\beta = \bar{x}_1 - \bar{x}_2 \bmod c$ , the  $\Delta$ -sum is bounded by

$$\ll \frac{P}{1 + \frac{P}{c} \|\beta\|} \cdot \mathcal{W}_{m_1, m_2},$$

with  $\mathcal{W}_{m_1, m_2}$  a smooth weight obeying  $\partial_{m_j}^i \mathcal{W} \ll X^{-i}$ . Hence

$$\mathcal{S} \ll \sum_{\substack{x_1, x_2 \bmod c \\ (x_i, c)=1}} \frac{P}{1 + \frac{P}{c} \|\bar{x}_1 - \bar{x}_2\|} \left| \sum_{m \asymp X} e\left(\frac{m(x_1 + \bar{x}_1 - x_2 - \bar{x}_2)}{c}\right) \mathcal{W}_m \right|^2.$$

*Completion in  $m$ .* By Poisson summation modulo  $c$ ,

$$\left| \sum_{m \asymp X} e\left(\frac{m\Theta}{c}\right) \mathcal{W}_m \right|^2 \ll X \left(1 + \frac{X}{c}\right),$$

uniformly in  $\Theta \bmod c$ .

*Sum over units.* Thus

$$\mathcal{S} \ll X \left(1 + \frac{X}{c}\right) \sum_{\substack{x_1, x_2 \bmod c \\ (x_i, c)=1}} \frac{P}{1 + \frac{P}{c} \|\bar{x}_1 - \bar{x}_2\|}.$$

The map  $x \mapsto \bar{x}$  permutes  $(\mathbb{Z}/c\mathbb{Z})^\times$ , so this equals

$$\phi(c) \sum_{\substack{y \bmod c \\ (y, c)=1}} \frac{P}{1 + \frac{P}{c} \|y\|}.$$

Bounding by the full sum over  $0 \leq y < c$  gives

$$\sum_{y=0}^{c-1} \frac{P}{1 + \frac{P}{c} \|y\|} \ll c + c \log(2 + P/c) \ll_\varepsilon (P + c) c^\varepsilon.$$

Therefore

$$\sum_{|\Delta| \leq P} |\Sigma_{q,r}(\Delta)|^2 \ll_\varepsilon X \left(1 + \frac{X}{c}\right) (P + c) c^{1+\varepsilon}.$$

*Final simplification.* Absorb  $1 + X/c \ll X^\varepsilon c^\varepsilon$  into the error. This yields the claimed bound

$$\sum_{|\Delta| \leq P} |\Sigma_{q,r}(\Delta)|^2 \ll_\varepsilon (P + c) c^{1+2\varepsilon} X^{1+2\varepsilon}. \quad \square$$

*Remark D.6* (Oldforms/Eisenstein and uniformity in  $q$ ). Lemma D.4 includes oldforms and Eisenstein; their geometric contributions have the same Kloosterman-Bessel shape with identical kernel bounds, so Lemma D.5 holds uniformly in the full spectrum. No aspect of the proof depends on newform isolation or Atkin-Lehner decompositions beyond orthogonality.

## Appendix I.5 Parameter box

For clarity we record the global parameter choices:

- Minor-arc cutoff:  $Q = N^{1/2-\varepsilon}$  with fixed  $\varepsilon \in (0, 10^{-2})$ .
- Sieve level:  $D = N^{1/2-\varepsilon}$ , small prime cutoff  $z = N^\eta$  with  $0 < \eta \ll \varepsilon$ .
- Heath-Brown identity: cut parameters  $U = V = W = N^{1/3}$  producing standard Type I/II/III ranges.
- Amplifier: primes in  $[P, 2P]$  with  $P = X^\vartheta$ ,  $0 < \vartheta < 1/6 - \kappa$ .
- Type III saving:  $\delta = \frac{1}{1000} \min\{\kappa, \frac{1}{2} - 3\vartheta\}$ .

## Appendix I.6 Auxiliary analytic inputs used in Part B

We record the external inputs used in Lemma B.2; full proofs are standard and can be found in the cited references.

**Lemma D.7** (Smooth Halász with divisor weights). *Let  $f$  be a completely multiplicative function with  $|f| \leq 1$ . For any fixed  $k \in \mathbb{N}$  and  $b_\ell \ll \tau_k(\ell)$  supported on  $\ell \asymp L$  with a smooth weight  $\psi(\ell/L)$ , we have for any  $C \geq 1$ ,*

$$\sum_{\ell \asymp L} b_\ell f(\ell) \psi(\ell/L) \ll_k L(\log L)^{-C}$$

*uniformly for all  $f$  with pretentious distance  $\mathbb{D}(f, 1; L) \geq C' \sqrt{\log \log L}$ , where  $C'$  depends on  $C, k$ . In particular the bound holds for  $f(n) = \lambda(n)\chi(n)$  when  $\chi$  is non-pretentious. References: Granville–Soundararajan (Pretentious multiplicative functions) and IK, §13; Harper (short intervals), with smoothing uniformity.*

**Lemma D.8** (Log-free exceptional-set count). *Fix  $C_1 \geq 1$ . For  $Q \leq L^{1/2}(\log L)^{-100}$ , the set*

$$\mathcal{E}_{\leq Q}(L; C_1) := \{\chi \pmod{q} : q \leq Q, \mathbb{D}(\lambda\chi, 1; L) \leq C_1\}$$

*has cardinality  $\#\mathcal{E}_{\leq Q}(L; C_1) \ll Q(\log(QL))^{-C_2}$  for some  $C_2 = C_2(C_1) > 0$ . This is a standard log-free zero-density consequence in pretentious form; see Montgomery–Vaughan, Ch. 12; Gallagher; IK, Thm. 12.2 and related log-free variants.*

**Lemma D.9** (Siegel-zero handling). *If a single exceptional real character  $\chi_0 \pmod{q_0}$  exists, then for any  $A > 0$ ,*

$$\sum_{\ell \asymp L} b_\ell \lambda(\ell) \chi_0(\ell) \psi(\ell/L) \ll L \exp(-c\sqrt{\log L})$$

*uniformly for  $b_\ell \ll \tau_k(\ell)$ , with an absolute  $c > 0$ . References: Davenport, Ch. 13; IK, §11 (Deuring–Heilbronn phenomenon).*

## Appendix I.7 Admissible parameter tuple and verification

We fix explicit values valid for large  $N$ :

$$\varepsilon = 10^{-3}, \quad \eta = 10^{-4}, \quad \kappa = 10^{-3}, \quad \vartheta = \kappa/8 = 1.25 \times 10^{-4}.$$

Then  $Q = N^{1/2-\varepsilon}$  and for Type II we have  $L \geq N^\eta$ , hence  $Q \leq L^{1/2}(\log L)^{-100}$  for large  $N$ , so Lemma D.8 applies. In Part C,  $P = X^\vartheta$  satisfies  $\vartheta < 1/6 - \kappa$ , and

$$\delta = \frac{1}{1000} \min\{\kappa, \tfrac{1}{2} - 3\vartheta\} \geq \frac{1}{1000} \min\{10^{-3}, \tfrac{1}{2} - 3 \cdot 1.25 \times 10^{-4}\} \geq 5 \times 10^{-7}.$$

Choose the log-power parameters  $A \geq 10$  and  $B = B(A, k, \eta)$  large (from Lemma B.2). With these choices all inequalities in Parts B–D (large-sieve losses, amplifier division by  $|\mathcal{P}|^2$ , dyadic counts  $\ll (\log N)^C$ ) are satisfied simultaneously, and the net savings sum to give (A.1).

## Appendix I.8 Deterministic balanced signs for the amplifier

**Lemma D.10** (Balanced signs). *Let  $\mathcal{P} = \{p \in [P, 2P] : p \text{ prime}\}$ . There exists a deterministic choice of signs  $\{\varepsilon_p\}_{p \in \mathcal{P}} \subset \{\pm 1\}$  with  $\sum_{p \in \mathcal{P}} \varepsilon_p = 0$ . Moreover, for every integer  $\Delta$ ,*

$$\left| \sum_{p \in \mathcal{P}} \varepsilon_p \varepsilon_{p+\Delta} \right| \leq \#\{p \in \mathcal{P} : p + \Delta \in \mathcal{P}\} \leq |\mathcal{P}| \cdot \mathbf{1}_{|\Delta| \leq 2P}.$$

*Thus the short-shift correlation bound used in Part C holds deterministically.*

*Proof.* Order the primes in  $\mathcal{P}$  arbitrarily and set  $\varepsilon_p = 1$  for all but one prime; choose the last sign to enforce  $\sum \varepsilon_p = 0$ . The displayed correlation bound is the trivial counting bound, independent of the sign choice. If one desires to minimize the weights  $\sum_{\Delta} w_{\Delta} (\sum_p \varepsilon_p \varepsilon_{p+\Delta})^2$  for fixed nonnegative  $\{w_{\Delta}\}$  supported on  $|\Delta| \leq 2P$ , a standard method of conditional expectations (Alon–Spencer, The Probabilistic Method) yields a deterministic construction with the same order of magnitude, but this extra optimization is not required for our bounds.  $\square$

## References (standard sources)

H. Iwaniec and E. Kowalski, *Analytic Number Theory*, AMS Colloquium Publications, Vol. 53. H. Montgomery and R. Vaughan, *Multiplicative Number Theory I. Classical Theory*, Cambridge Univ. Press. H. Davenport, *Multiplicative Number Theory*, 3rd ed., Springer. J.-M. Deshouillers and H. Iwaniec, Kloosterman sums and Fourier coefficients of cusp forms, *Ann. Inst. Fourier* (1982). A. Granville and K. Soundararajan, Pretentious multiplicative functions and analytic number theory (various papers/notes). A. Harper, Bounds for multiplicative functions in short intervals. N. Alon and J. Spencer, *The Probabilistic Method* (for conditional expectations derandomization).