I shall begin by describing the most general context in which the elementary approach to the prime number theorem works (at least at present). We consider the case of Beurlings generalized integers: Let us have a set of real numbers p_i, $i = 1, 2, 3, \cdots$

$$1 < p_1 \le p_2 \le \cdots \le p_i \le \cdots$$

such that $p_i \to \infty$ as $i \to \infty$, we form all possible finite products

$$\prod p_i^{\alpha_i}$$

and order them according to magnitude or size

$$n_1 = 1, \; n_2 = p_1, \cdots, \; n_i \le n_{i+1}, \to \infty,$$

We denote by $N(x)$ the number of $n_i \le x$ and assume we have an asymptotic law

$$N(x) = Ax + R(x),$$

where we shall assume $A=1$ and

$$R(x) = O\left(\frac{x}{\log^2 x}\right),$$

and shall try to establish the P.N.T, that is, if we denote by $\pi(x)$ the number of $p_i \le x$

We shall show that

$$\pi(x) = \frac{x}{\log x} + o\left(\frac{x}{\log x}\right),$$

in the form

$$\vartheta(x) = x + o(x),$$

where $\vartheta(x) = \sum_{p_i \le x} \log p_i$,

Similarly we use the notation

$$\psi(x) = \sum_{p_i^\alpha \le x} \log p_i .$$

and

$$\Lambda(n_i) = \begin{cases} \log p_i & \text{if } n_i = p_i^\alpha, \alpha > 0, \\ 0 & \text{otherwise}. \end{cases}$$

For simplicity, I shall drop the indices and write n, m or d for the generalized integers, p, q and r for generalized primes, and use greek letter μ, ν to denote ordinary integers, constants will be denoted by capital latin letters. x, y, t, u, v real variables

We define $\mu(n) = (-1)^{\nu}$ if n is the product of ν distinct primes p and otherwise $\mu(n) = 0$. We also write

$$d \mid n \quad \text{if} \quad n = \prod p_i^{\alpha_i},$$

$$d = \prod p_i^{\beta_i} \text{ and } \alpha_i \geq \beta_i \text{ for all } i.$$

We have then $\displaystyle\sum_{d \mid n} \mu(d) = \begin{cases} 0 & \text{for } n \neq 1, \\ 1 & \text{for } n = 1. \end{cases}$

We need some preliminary estimations, we first shall show

$$\sum_{d \leq x} \frac{\mu(d)}{d} = O(1).$$

We have for $x \geq 1$,

$$1 = \sum_{n \leq x} \sum_{d \mid n} \mu(d) = \sum_{d \leq x} \mu(d) N\left(\frac{x}{d}\right) =$$

$$= x \sum_{d \leq x} \frac{\mu(d)}{d} + O\left(\sum_{d \leq x} \left| R\left(\frac{x}{d}\right) \right| \right).$$

Here

$$\sum_{d \leq x} \left| R\left(\frac{x}{d}\right) \right| = x \sum_{d \leq x} \frac{\varepsilon\left(\frac{x}{d}\right)}{d\left(1 + \log \frac{x}{d}\right)^2},$$

where we here and in the future denote by $\varepsilon(x)$ a function that tends to zero as $x \to \infty$. Dividing the interval $(1, x]$ into subintervals $xe^{-\nu} \le d \le xe^{1-\nu}$ for $\nu = 1, 2, \ldots, \log x$, we get

$$\sum_{d \le x} \frac{\varepsilon\left(\frac{x}{d}\right)}{d\left(1+\log\frac{x}{d}\right)^2} < \sum_{1 \le \nu \le \log x} \frac{\varepsilon(e^{\nu-1})\, N\left(\frac{x}{e^{\nu-1}}\right)}{xe^{-\nu}\left(1+\nu-1\right)^2}$$

$$< A \sum_{\nu \ge 1} \frac{\varepsilon(e^{\nu-1})}{\nu^2} = O(1),$$

Thus

$$1 = x \sum_{d \le x} \frac{\mu(d)}{d} + O(x),$$

and

$$\sum_{d \le x} \frac{\mu(d)}{d} = O(1),$$

follows.

We next show

$$\sum_{d \le x} \frac{\mu(d)}{d} \log \frac{x}{d} = o(\log\log x),$$

consider first

$$\sum_{n \le y} \frac{1}{n} = \int_1^y \frac{1}{t}\, dN(t) =$$

$$= \frac{N(y)}{y} + \int_1^y \frac{N(t)}{t^2}\,dt = 1 + o\left(\frac{1}{\log^2 y}\right)$$

$$+ \log y + \int_1^y \frac{\varepsilon(t)}{t\,(1+\log t)^2}\,dt$$

$$= \log y + 1 + o\left(\frac{1}{\log^2 y}\right) + \int_1^\infty \frac{\varepsilon(t)\,dt}{t(1+\log t)^2}$$

$$+ o\left(\int_y^\infty \frac{dt}{t(1+\log t)^2}\right)$$

$$= \log y + C + o\left(\frac{1}{1+\log y}\right).$$

So we can write

$$\log\frac{x}{d} = \sum_{n\le \frac{x}{d}}\frac{1}{n} - C + \frac{\varepsilon(\frac{x}{d})}{1+\log\frac{x}{d}},$$

and so

$$\sum_{d\le x}\frac{\mu(d)}{d}\log\frac{x}{d} = \sum_{d\le x}\frac{\mu(d)}{d}\sum_{n\le \frac{x}{d}}\frac{1}{n}$$

$$- C\sum_{d\le x}\frac{\mu(d)}{d} + \left(\sum_{d\le x}\frac{\varepsilon(\frac{x}{d})}{d(1+\log\frac{x}{d})}\right),$$

here

$$\sum_{d\le x}\frac{\mu(d)}{d}\sum_{n\le \frac{x}{d}}\frac{1}{n} = \sum_{nd\le x}\frac{\mu(d)}{nd} =$$

$$= \sum_{m\le x}\frac{1}{m}\sum_{d\mid m}\mu(d) = 1,$$

also
$$\sum_{d\le x} \frac{\mu(d)}{d} = O(1),$$

and
$$\sum_{d\le x} \frac{\varepsilon(\frac{x}{d})}{d(1+\log\frac{x}{d})} <$$

$$< \sum_{1\le\nu\le\log x} \frac{\varepsilon(e^{\nu-1})\, e^{\nu} N(xe^{1-\nu})}{x\ \ (1+\nu-1)}$$

$$< A \sum_{1\le\nu\le\log x} \frac{\varepsilon(e^{\nu-1})}{\nu} = o(\log\log x).$$

This gives
$$\sum_{d\le x} \frac{\mu(d)}{d} \log\frac{x}{d} = o(\log\log x).$$

We finally show that
$$\sum_{d\le x} \frac{\mu(d)}{d} \log^2\frac{x}{d} = \log x + \sum_{n\le x} \frac{\Lambda(n)}{n} + o(\log x)$$

$$= \log x + \sum_{p\le x} \frac{\log p}{p} + o(\log x).$$

We have
$$\sum_{d\le x} \frac{\mu(d)}{d} \log^2\frac{x}{d} = \sum_{d\le x} \frac{\mu(d)}{d} \log\frac{x}{d} \sum_{n\le\frac{x}{d}} \frac{1}{n}$$

$$- C \sum_{d\le x} \frac{\mu(d)}{d} \log\frac{x}{d} + O\Big(\sum_{d\le x} \frac{\varepsilon(\frac{x}{d})\log\frac{x}{d}}{d(1+\log\frac{x}{d})}\Big).$$

The first term on the righthand side equals $\sum_{md \le x} \frac{\mu(d) \log \frac{x}{d}}{md} =$

$$= \sum_{m \le x} \frac{1}{m} \sum_{d|m} \mu(d) \log \frac{x}{d},$$

but $\sum_{d|m} \mu(d) \log \frac{x}{d} = \begin{cases} \log x, & \text{for } m = 1 \\ \Lambda(m), & \text{for } m > 1. \end{cases}$

So $\sum_{d \le x} \frac{\mu(d)}{d} \log \frac{x}{d} \sum_{m \le \frac{x}{d}} \frac{1}{m} =$

$$= \log x + \sum_{m \le x} \frac{\Lambda(m)}{m} = \log x + \sum_{p \le x} \frac{\log p}{p} + O(1).$$

Also $\sum_{d \le x} \frac{\varepsilon(\frac{x}{d}) \log \frac{x}{d}}{d(1 + \log \frac{x}{d})} < \sum_{d \le x} \frac{\varepsilon(\frac{x}{d})}{d} <$

$$< \sum_{1 \le \nu \le \log x} \frac{\varepsilon(e^{\nu-1}) e^{\nu}}{x} N\left(\frac{x}{e^{\nu-1}}\right)$$

$$< A \sum_{1 \le \nu \le \log x} \varepsilon(e^{\nu-1}) = o(\log x).$$

Thus altogether we get:

$$\sum_{d\le x} \frac{\mu(d)}{d} \log^2 \frac{x}{d} = \log x + \sum_{p\le x} \frac{\log p}{p}$$

$$+ o(\log x).$$

Now consider

$$\sum_{n\le x} \sum_{d/n} \mu(d) \log^2 \frac{x}{d} =$$

$$\sum_{d\le x} \mu(d) \log^2 \frac{x}{d} \cdot N\left(\frac{x}{d}\right) =$$

$$x \sum_{d\le x} \frac{\mu(d)}{d} \log^2 \frac{x}{d} + O\left(x \sum_{d\le x} \frac{\varepsilon(\frac{x}{d}) \log^2 \frac{x}{d}}{d\,(1+\log\frac{x}{d})^2}\right).$$

The last term is

$$O\left(x \sum_{d\le x} \frac{\varepsilon(\frac{x}{d})}{d} \right) = o(x \log x),$$

so we get altogether

$$\sum_{n\le x} \sum_{d/n} \mu(d) \log^2 \frac{x}{d} = x \log x$$

$$+ x \sum_{p\le x} \frac{\log p}{p} + o(x \log x).$$

On the other hand, it is easily

verifyed that for $n \leq x$

$$\sum_{d|n} \mu(d) \log^2 \frac{x}{d} = \begin{cases} \log^2 x, & \text{for } n=1 \\ \log \frac{x^2}{p} \log p, & \text{for } n = p^\nu, \nu>0, \\ 2 \log p \log q, & \text{for } n = p^\nu q^\mu, p \neq q. \\ 0 & \text{if } n \text{ has 3 or more prime factors} \end{cases}$$

Seen either by induction, or by noting that it is the second derivative of

$$x^y \sum_{d|n} \mu(d) d^{-y} = x^y \prod_{p|n} (1-p^{-y})$$

at $y=0$.

So we have also

$$\sum_{n \leq x} \sum_{d|n} \mu(d) \log^2 \frac{x}{d} = \log^2 x$$

$$+ \sum_{p^\nu \leq x} \log \frac{x^2}{p} \log p + \sum_{\substack{p^\nu q^\mu \leq x \\ p \neq q}} \log p \log q .$$

Comparing results we get

$$\sum_{p \leq x} \log \frac{x^2}{p} \log p + \sum_{pq \leq x} \log p \log q$$

$$= x \log x + x \sum_{p \leq x} \frac{\log p}{p} + o(x \log x).$$

We still need to estimate the sum

$$\sum_{p \le x} \frac{\log p}{p} .$$

We consider the sum

$$\sum_{n \le x} \log n = \int_1^x \log t \, dN(t) =$$

$$= N(x) \log x - \int_1^x \frac{N(t)}{t} dt$$

$$= x(\log x - 1) + o\left(\frac{x}{\log x}\right).$$

Also since

$$\log n = \sum_{d|n} \Lambda(d) \quad \text{we have,}$$

using $N(x) > \frac{1}{A} x$ for $x \geq 1$,

that

$$\sum_{n \le x} \log n = \sum_{d \le x} \Lambda(d) N\left(\frac{x}{d}\right) >$$

$$> \frac{1}{A} x \sum_{d \le x} \frac{\Lambda(d)}{d} .$$

Comparing this with the result above we get at first

$$\sum_{d \le x} \frac{\Lambda(d)}{d} = O(\log x),$$

Inserting this bound in our earlier inequality we get at first

$$\vartheta(x) = \sum_{p \le x} \log p = O(x), \text{ or}$$

$$\psi(x) = \sum_{d \le x} \Lambda(d) = O(x).$$

We use this to estimate

$$\sum_{d \le x} \Lambda(d)\, N\left(\frac{x}{d}\right) \text{ better,}$$

We get

$$\sum_{d \le x} \Lambda(d)\, N\left(\frac{x}{d}\right) = x \sum_{d \le x} \frac{\Lambda(d)}{d} +$$

$$+ x \sum_{d \le x} \frac{\Lambda(d)}{d} \frac{\varepsilon\left(\frac{x}{d}\right)}{\left(1 + \log \frac{x}{d}\right)^2}.$$

Here

$$x \sum_{d \le x} \frac{\Lambda(d)}{d} \frac{\varepsilon\left(\frac{x}{d}\right)}{\left(1 + \log \frac{x}{d}\right)^2} \le$$

$$\le \sum_{1 \le \nu \le \log x} \psi(x e^{1-\nu}) \frac{e^{\nu}\, \varepsilon(e^{\nu-1})}{\nu^2} <$$

$$< A x \sum_{\nu \ge 1} \frac{\varepsilon(e^{\nu-1})}{\nu^2} = O(x).$$

Thus

$$\sum_{d \le x} \Lambda(d)\, N\left(\frac{x}{d}\right) = x \sum_{d \le x} \frac{\Lambda(d)}{d} + O(x),$$

but since also

$$\sum_{d \le x} \Lambda(d)\, N\left(\frac{x}{d}\right) = x(\log x - 1) + o\left(\frac{x}{\log x}\right)$$

we get by comparison

$$\sum_{d \le x} \frac{\Lambda(d)}{d} = \log x + O(1)$$

or $$\sum_{p \le x} \frac{\log p}{p} = \log x + O(1),$$

and so

$$\sum_{p \le x} \log \frac{x^2}{p} \log p + \sum_{pq \le x} \log p \log q =$$
$$= 2x \log x + o(x \log x).$$

Since from $\vartheta(x) = O(x)$

$$\sum_{p \le x} \log \frac{x}{p} \log p = \int_1^x \log \frac{x}{t}\, d\vartheta(t)$$
$$= \int_1^x \frac{\vartheta(t)}{t}\, dt = O(x),$$

we may rewrite the asymptotic formula above as

$$\sum_{p \le x} \log^2 p + \sum_{pq \le x} \log p \log q = 2x \log x + o(x \log x)$$

or $$\log x\, \vartheta(x) + \sum_{p \le x} \log p\, \vartheta\left(\frac{x}{p}\right) = 2x \log x + o(x \log x).$$

We also can rewrite $\sum_{p\leq} \frac{\log p}{p} = \log x + O(1)$

as $$\int_1^x \frac{\vartheta(t)}{t^2}\,dt = \log x + O(1)$$

By partial summation we get from

$$\sum_{p\leq x} \log^2 p + \sum_{pq\leq x} \log p \log q = 2x\log x + o(x\log x)$$

that

$$\sum_{p\leq x} \log p + \sum_{pq\leq x} \frac{\log p \log q}{\log pq} = 2x + o(x).$$

From this

$$\sum_{pq\leq x} \log p \log q = \sum_{p\leq x} \log p \sum_{q\leq \frac{x}{p}} \log q =$$

$$= 2x\sum_{p\leq x} \frac{\log p}{p} - \sum_{p\leq x} \log p \sum_{qr\leq x} \frac{\log q \log r}{\log qr} +$$

$$+ O\Big(x\sum_{p\leq x} \varepsilon\big(\tfrac{x}{p}\big) \frac{\log p}{p}\Big) = 2x\log x$$

$$- \sum_{qr\leq x} \frac{\log q \log r}{\log qr}\, \vartheta\Big(\frac{x}{qr}\Big) + o(x\log x).$$

Combining this with

$$\log x\, \vartheta(x) + \sum_{pq\leq x} \log p \log q = 2x\,\cancel{\log}\, x + o(x\log x)$$

We get

$$\log x\, \vartheta(x) = \sum_{pq \le x} \frac{\log p \log q}{\log pq}\, \vartheta\left(\frac{x}{pq}\right) + o(x \log x).$$

If we write $\vartheta(x) = x + \rho(x)$ the two last equations give

$$\log x\, \rho(x) = -\sum_{p \le x} \log p\, \rho\left(\frac{x}{p}\right) + o(x \log x)$$

and

$$\log x\, \rho(x) = \sum_{pq \le x} \frac{\log p \log q}{\log pq}\, \rho\left(\frac{x}{pq}\right) + o(x \log x).$$

From this

$$2 \log x\, |\rho(x)| \le \sum_{p \le x} \log p\, \left|\rho\left(\frac{x}{p}\right)\right| + \sum_{pq \le x} \frac{\log p \log q}{\log pq} \left|\rho\left(\frac{x}{pq}\right)\right| + o(x \log x).$$

If we write

$$2s(y) = \sum_{p \le y} \log p + \sum_{pq \le y} \frac{\log p \log q}{\log pq} = 2y + o(y),$$

we may rewrite the above inequality

as
$$|\rho(x)| \le \frac{1}{\log x}\int_1^x |\rho(\tfrac{x}{t})|\,ds(t) + O(x).$$
Let us now compare
$$\int_1^x |\rho(\tfrac{x}{t})|\,ds(t) \text{ and } \int_1^x |\rho(\tfrac{x}{t})|\,dt,$$
in the form
$$\int_0^{\log x} |\rho(xe^{-u})|\,ds(e^u) \text{ and } \int_0^{\log x} |\rho(xe^{-u})|e^u\,du.$$
Let δ be a small positive number, then exists then a $u_0(\delta)$, such that for $u \ge u_0(\delta)$
$$|s(e^{u+\delta}) - s(e^u) - e^u(e^\delta - 1)| < \delta^2 e^u,$$
and such that also for $u \ge u_0(\delta)$
$$|\rho(e^{u'}) - \rho(e^{u''})| \le |e^{u'} - e^{u''}| + \delta^2 e^u,$$
if both u' and u'' lie in the interval $(u, u+\delta)$.

We now divide the interval $(0, \log x]$ into $N = [\frac{\log x}{\delta}] + 1$ intervals of equal length δ except possibly the last.
We now look at an interval $(\nu\delta, (\nu+1)\delta)$ when $\nu\delta \ge u_0(\delta)$, and

$\log x - (\nu+1)\delta \geq u_0(\delta)$.

We have

$$\mathcal{J}_\nu = \int_{\nu\delta}^{(\nu+1)\delta} |\rho(xe^{-u})| e^{u} du = |\rho(xe^{-u'})| e^{u'} \delta,$$

where $\nu\delta \leq u' \leq (\nu+1)\delta$, and

$$J_\nu = \int_{\nu\delta}^{(\nu+1)\delta} |\rho(xe^{-u})| \, d\, s(e^{u}) = |\rho(xe^{-u''})| \left(s(e^{(\nu+1)\delta}) - s(e^{\nu\delta}) \right),$$

where $\nu\delta \leq u'' \leq (\nu+1)\delta$.

Using now the inequalities

$$|\rho(xe^{-u'})| - |\rho(xe^{-u''})| \leq |\rho(xe^{-u'}) - \rho(xe^{-u''})| \leq$$

$$\leq x|e^{-u'} - e^{-u''}| + \delta^2 x e^{-\nu\delta} \leq$$

$$\leq x e^{-\nu\delta}\delta + \delta^2 x e^{-\nu\delta},$$

and

$$|s(e^{(\nu+1)\delta}) - s(e^{\nu\delta}) - e^{\nu\delta}(e^{\delta}-1)| \leq \delta^2 e^{\nu\delta},$$

which implies

$$|s(e^{(\nu+1)\delta}) - s(e^{\nu\delta}) - \delta e^{\nu\delta}| \leq 2\delta^2 e^{\nu\delta},$$

we see that the difference between $|\mathcal{J}_\nu - J_\nu| \leq 5x\delta^2$.

Since there are $< \frac{\log x}{\delta}$ intervals in $(u_0(\delta), \log x - u_0(\delta))$ and the parts from $(0, u_0(\delta))$ and $(\log x - u_0(\delta), \log x)$ contribute only an amount $\leq c(\delta)x$ to the two integrals, we get

$$\left| \int_0^{\log x} |\rho(xe^{-u})| \, ds(e^u) - \int_0^{\log x} |\rho(xe^{-u})| e^u \, du \right| \leq$$

$$\leq 5\delta x \log x + c(\delta)x.$$

Since δ can be chosen arbitrarily close to zero, we see that the difference is $o(x\log x)$. Thus we have the inequality

$$|\rho(x)| \leq \frac{1}{\log x} \int_0^{\log x} |\rho(xe^{-u})| e^u \, du + o(x),$$

or writing u for $\log x - u$, we get

$$|\rho(x)| \leq \frac{x}{\log x} \int_0^{\log x} |\rho(e^u)| e^{-u} \, du + o(x).$$

Our earlier equation

$$\sum_{p \leq x} \frac{\log p}{p} = \log x + O(1),$$

may be rewritten as:

$$\int_1^x \frac{\vartheta(t)}{t^2}\,dt = \log x + O(1),$$

or

$$\int_1^x \frac{\rho(t)}{t^2}\,dt = O(1)$$

and again with some positive constant $\mathscr{C}$

$$\left| \int_0^{\log x} \rho(e^u)e^{-u}\,du \right| \le \mathscr{C},$$

or

$$\left| \int_u^{u+\lambda} \rho(e^u)e^{-u} \right| \le 2\mathscr{C}.$$

Assume now that $|\rho(x)| \le \alpha x + o(x)$, where $\alpha \le 1$. Choose λ so that

$$\lambda\alpha = 2\mathscr{C} + \tfrac{3}{10}\alpha^2.$$

Then either

$$\int_u^{u+\lambda} |\rho(e^u)|e^{-u}\,du \le \lambda\alpha - \tfrac{3}{10}\alpha^2$$

or there is a u' in $(u, u+\lambda)$ where $\rho(e^u)$ changes sign. In the latter case

$$|\rho(e^{u'+h})| \le |e^{u'+h} - e^{u'}| + o(e^{u'+|h|}),$$

and we get an interval on each side

of u' where

$$|\rho(e^{u'+h})|e^{-u'-h} < \alpha,$$

one of these lies entirely in $[u, u+\lambda)$ and one finds its contribution to the integral is small enough to make

$$\int_u^{u+\lambda} |\rho(e^u)|e^{-u}\,du \le \lambda\alpha - (1-\log 2)\alpha^2 + o(1)$$

$$\le \lambda\alpha - \frac{3}{10}\alpha^2, \text{ for } u \ge u_0.$$

Our inequality

$$|\rho(x)| \le \frac{x}{\log x}\int_0^{\log x} |\rho(e^u)|e^{-u}\,du + o(x),$$

now gives

$$|\rho(x)| \le \left(\alpha - \frac{3\alpha^2}{10\lambda}\right)x + o(x) =$$

$$= \left(\alpha - \frac{3\alpha^3}{20\mathscr{C} + 3\alpha^2}\right)x + o(x).$$

The iteration

$$\alpha_{n+1} = \alpha_n\left(1 - \frac{3\alpha_n^2}{20\mathscr{C} + 3\alpha_n^2}\right),$$

is easily seen to converge to zero, so we get

$$\rho(x) = o(x),$$

from which

$$\vartheta(x) = x + o(x),$$

and

$$\pi(x) = \frac{x}{\log x} + o\left(\frac{x}{\log x}\right)$$

follows immediately.

The later developments of the elementary proof of P.N.T. have gone in the direction of getting a better and better remainder term.

The main developments: Bombieri and Wirsing proved independently (and using rather different methods) that (we are now talking of the natural integers):

$$\pi(x) = \int_2^x \frac{dt}{\log t} + O\left(\frac{x}{\log^k x}\right),$$

for k an arbitrary large constant. Later this has by Diamond and others been improved so that the remainder term is $O(x e^{-(\log x)^{\alpha}})$ for some small positive constant.