Information Extraction in Active Time/Frequency Based Process Measurement

T. A. Groover ℜℑ−

Abstract

In applications employing energy pulses for measurement, uncertainty is inevitable as in all other measurement scenarios. When inaccuracies due to all controllable parameters are accounted for, there remains measurement uncertainty due to random noise. This paper is directed towards showing that an energy pulse in the presence of noise can be characterized by a self-information measure. This measure can be useful in characterizing time or doppler shift demarcation and corresponds to symbol discrimination events in electrical communications and indirectly shown to be their analog via convergent characteristics. A benefit arising from the effort is a resulting novel set of formulas for the time-of-arrival resolution which takes, as input, signal and noise spectral profiles. And these can be extended to apply to frequency-at-arrival resolution by an ambiguity function application. The methods employed can also open a path for a quantification of Shannon entropy for a single time-of-arrival measurement.

I. Introduction

Estimation and extraction of information from a process have elements common to branches of engineering such as classical and discrete signal processing and statistical communications. As signal processing is common to electrical communications, pulsed sonar and radar, lidar, and transit-time ultrasonic flow measurement, an additional level of effort is paid to classical information theory for insights which can be extended to estimation theory in general as shown in this section. Probably the first effort to apply information theory to time measurement was by Woodward[1] and Davies[2] in the 1950's. In the discussion here theory will be applied to the problem of time measurement resolution as it relates to the central problem of ultrasonic flow measurement: that of measurement of signal transit-time, transit-time-difference, or more simply, arrival point in time discrimination. This problem is in common to one in the radar and sonar fields as it relates to range measurement, and the related problem of Doppler shift measurement. The impetus for the effort here began with a desire for an expression for time or range resolution (i.e. the standard deviation of an ensemble or cluster of

1 Woodward, P. M., "Probability and Information Theory with Applications to Radar" Pergamon Press, Oxford, 1953
2 Woodward, P. M. and Davies, I. L., "A Theory of Radar Information" IRE Trans. Inf. Theory, vol. 1, pp. 108–113, 1953

repetitive measurement data) as a function of both signal and noise amplitude spectra. The desired pathway to such an expression is a direct if somewhat heuristic application of the Shannon-Hartley channel capacity theorem. As the title suggests, there is an extensive reliance on a central abstraction – the conversion of a problem in classical signals and systems into a problem of discrete events (termed *timing events*) constructed as threshold crossings of a complex signal trajectory. As such the thresholds are entirely abstract, but are a part of an approach to fashion a bridge between the discrete symbols of communication to the continuous realm of measurement, so as to apply the concepts of *self information* and the related *entropy* as arises in information theory. An important result is that a measurement is characterized with a self information value and in an ensemble of similar estimations, an entropy value. This result from Part Two will be previewed and generalized in Part One immediately following. And then the results from Part One can be applied to replicate measurements in the realm of Part Two.

The ideas in this paper arise from the author's experience with ultrasonic flow measurement. However terminology from that field will be de-emphasized in favor of terminology common in science and engineering, and specific shades of meaning have been employed in some cases for the specific usage here. And in a few cases terms have been devised. For clarity with application of these meanings, Appendix I provides a glossary of terms. In the cases of devised terms the glossary definition of these devices may add to the text body definition. And the endeavor will address topics that have been thoroughly covered over a period of decades, most extensively in the radar literature, and also from sonar development. This writer, not experienced in those fields, will possibly employ methods and terminology of unconventional nature as seen by those in the fields of radar and sonar. The use of estimator variables, employed in Part One, will be eschewed in subsequent sections. The signal naming may seem unconventional but are distinguished primarily as subscripts, as there are varying forms of the conveyed information being borne at different stages of the analytic treatment.

Part One.

II. Measurement Theory, Surprisal, Self Information, and Entropy

Over the decades there have been proposals for calculation of self-information for a continuous random variable (starting with Shannon, then Fisher) with the $-p\log(p)$ incorporated into an integrand expression (the so-called cumulative probability of continuous random variables) having no direct logical association with the discrete random variable considered in this section. Application of the $-p\log(p)$ expression to cumulative probability of continuous random variables yields 'differential entropy' which has no application for the present endeavor, and in any case is difficult to interpret as there can be probability distributions yielding negative entropy.

What will be attempted is a calculation which can be applied in a measurement scenario to individual data points. Although current practice is replete with entropy estimation of parameters of data, given unknown or arbitrary population probability density functions (PDF's) of measured parameters[3-4], the immediate exercise will be a straightforward one, to determine the self-information of an isolated data point and the related Shannon entropy as a result of a measurement (and this is in contradistinction to the Fisher information). There is seemingly no rigorous solution for this endeavor, consequently in this section there is significant heuristic and logical dependency.

Consider an estimator $\hat{A}$ of nonzero, scalar measurand A:

$$A, \hat{A} \in \mathbb{R}_{\neq 0}$$

Replicate measurements of an instance of A (termed also *estimations* yielding *data points*) can be grouped in sets termed *clusters*[5] symbolized as Q_{u} where the subscript 'u' denotes an unabridged or untruncated cluster. There is a cluster estimator function ω_A so that

$$\omega_A : \mathbb{R}_{\neq 0} \to Q_{\mathrm{u}} \subset \mathbb{R}_{\neq 0} \quad (1)$$

3 J. Beirlant, E. Dudewicz, L. Gyor£, and E. van der Meulen, "Nonparametric entropy estimation: An overview," *International Journal of Mathematical and Statistical Sciences*, vol. 6, no. 1, pp. 17–39, 1997.

4 Erik G. Miller, "A New Class Of Entropy Estimators For Multi-dimensional Densities", I.E.E.E International Conference on Acoustics, Speech, and Signal Processing, 2003

5 J. Martin Bland & Douglas G. Altman "Agreement Between Methods of Measurement with Multiple Observations Per Individual" J. Biopharmacological Stat 17:4, 2007, pp. 571-582

Then the measurement replication in time can be denoted with a discrete time argument t_j, random variable argument ϵ, and A as a parameter; and consider estimator bias B_u as argument. So the output is the estimator $\hat{A}_u$ of measurand A populating an untruncated cluster Q_u:

$$\omega_A(t_j, \epsilon, B_{\hat{A}}(t) | A) = \hat{A}_u \quad (2)$$

Since ω_A has a stochastic nature as measurement error input ϵ, then clearly estimator $\hat{A}_u$ is a random variable. The variable t_j will be a useful for possibly interchangeability of measuring methods or instruments having bias B_u. Thus B_u may be constant or time dependent.

There is a cluster operation ω_t which implements truncation, applying bounds $\{b_0, b_1\}$ for generating a truncated cluster Q_t, as

$$\omega_t : Q_u \big| b_0, b_1 \rightarrow Q_t \subset \left\{ \left[b_0, b_1 \right] \cup \left[-b_1, -b_0 \right] \right\} \subset \mathbb{R}_{\neq 0} \quad (3)$$

and executed by a party X, who determines the bounds for passing the data in Q_t to party Y. An unanswered question for now is whether ω_t fits the *function* proper definition but since ω_t^{-1} is an injective/non-surjective function (of doubtful utility), then a function interpretation of (3) would provide a truncated estimator $\hat{A}_t$ of measurand A:

$$\omega_t(\hat{A}_u | b_0, b_1) = \hat{A}_t \quad (4)$$

And to specify an untruncated Q_u as the source for a truncated Q_t ; the following is convenient for later reference, but also assume a requirement that each member of Q_u be processed by (4), with:

$$Q_t \subset Q_u \quad (5)$$

And so A will refer to a measurand value with related estimators $\hat{A}$, $\hat{A}_u$ and $\hat{A}_t$. and for convenience the two latter estimators will be from finite sets with cardinalities:

$$|Q_t| \leq |Q_u| < \infty \quad (6)$$

although it will be necessary for $|Q_u|$ to grow without bound in one step of the progress. Similar to the analysis in Part Two for signal arrival time measurement, $\hat{A}_u$ will be characterized with resolution limited by the random summed error component ϵ, with PDF f_ε (possibly unknown) and s.d. s_ε, and in the case of $\hat{A}_u$ resolution also *possibly* affected by possibly nonzero bias B_u. The quantity ϵ is as such an unobservable stochastic process except when A and B_u are known, referred to as a special case in any epistemic logic discussion. Then there is $\hat{A}$, unbiased and 'deterministic', relegated to the role of pure abstraction. The function f_ε although possibly unknown will be considered as having prior information attributes, that is of such there can be some knowledge based on Q_u possessed by party X (In the case of X omniscient of $\{A, B_u\}$ then X would have f_ε knowledge approaching ideal as Q_u grows very large.)

Assume for now that there is a non-Gaussian component of f_ε, for example as due to a mechanical vibration. So study of Q_u can be a source of prior information when addressing f_ε attributes and additionally those of Q_t. One of these prior information attributes will constitute a case to be explored here and that is that in this case its PDF exhibits the non-Gaussian component as a rectangular or quasi-rectangular profile due to mechanical vibration with a limit cycle. The truncation interval applied to obtain Q_t is based upon any prior information, including for example knowledge of Q_u, f_ε, the measurement method and/or the measurand A. Prior information is most often difficult to quantify and invariably incorporates probabilities or rather their estimates[6]. Fortunately in this Part 1, prior probabilities will play a logical role rather than a quantitative one. So from Eq. (3), the discussion is predicated on truncated set Q_t as:

$$0 < b_0 < |\hat{A}_t| < b_1 \quad (7)$$

And there will be a membership rule such that all of the members of Q_u which satisfy the requirements for members of Q_t are such members, expressed as set cardinality in conjunction with expression (5):

$$|Q_t| = \Pr\left[b_0 < |\hat{A}_u| < b_1\right] \times |Q_u| \quad (8)$$

6 Edwin T. Janes, "Prior Probabilities" IEEE Trans. Systems Science and Cybernetics, vol. sec-4, no. 3, 1968, pp. 227-241

Suppose the so-called "true value" (as from measurement literature[7]) A is unknowable (in epistemic logic referenced as its *knowability* attribute), there would be a conceptual substitution for it in some cases from the law of large numbers:

$$\mathrm{E}(\hat{A}_{\mathrm{u}} \,\big|\, |Q_{\mathrm{u}}| \to \infty) - B_{\mathrm{u}} = A \quad (9)$$

Now consider that the stochastic process may be isolated, expressed as a PDF of ϵ, given as f_{ε}. So in these cases the PDF of $\hat{A}_{\mathrm{u}}$ designated say $f_{\hat{a}_{\mathrm{u}}}(\hat{a}_{\mathrm{u}})$ can be:

$$f_{\hat{a}_{\mathrm{u}}}(\hat{a}_{\mathrm{u}}) = f_{\varepsilon}(\hat{a}_{\mathrm{u}} - A - B_{\mathrm{u}}) \quad (10)$$

that is, if A and B_{u} are knowable; if not then (10) becomes, from (9):

$$f_{\hat{a}_{\mathrm{u}}}(\hat{a}_{\mathrm{u}}) = f_{\varepsilon}\left[\hat{a}_{\mathrm{u}} - \mathrm{E}(\hat{A}_{\mathrm{u}} \,\big|\, |Q_{\mathrm{u}}| \to \infty)\right]$$

From this discussion note also that in the truncated case, a derived associated PDF can be expressed, employing the Heaviside step function $u()$. So the PDF of truncated estimator $\hat{A}_{\mathrm{t}}$, say termed $f_{\hat{a}_{\mathrm{t}}}(\hat{a}_{\mathrm{t}})$, can then be in terms of rectangular f_{ε}, based on $f_{\hat{a}_{\mathrm{u}}}(\hat{a}_{\mathrm{u}})$ from (10):

$$f_{\hat{a}_{\mathrm{t}}}(\hat{a}_{\mathrm{t}}) = \frac{f_{\hat{a}_{\mathrm{u}}}(\hat{a}_{\mathrm{t}})}{\Pr(\hat{A}_{\mathrm{u}} \in Q_t)} \times \begin{cases} u(\hat{a}_{\mathrm{t}} - b_0) - u(\hat{a}_{\mathrm{t}} - b_1) & \text{if } A > 0 \\ u(b_1 - \hat{a}_{\mathrm{t}}) - u(b_0 - \hat{a}_{\mathrm{t}}) & \text{if } A < 0 \end{cases} \quad (11)$$

the denominator being normalization. Now as noted, prior knowledge of f_{ε} may be that it is nearly rectangular as possibly determined from Q_{u} by inspection, so this will be presumed and useful, and to formalize the rectangular approximation of Eq. (11):

$$f_{\hat{a}_{\mathrm{t}}}(\hat{a}_{\mathrm{t}}) \approx \begin{cases} U(b_0, b_1) & \text{if } A > 0 \\ U(-b_1, -b_0) & \text{if } A < 0 \end{cases} \quad (12)$$

7 Ehrlich, C., Dybkaer, R. & Wöger, W. "Evolution of philosophy and description of measurement" Accred Qual Assur (2007) 12: 201. https://doi.org/10.1007/s00769-007-0259-4

where $U(b_0, b_1)$ is the uniform probability distribution supported over $[b_0, b_1]$. Clearly the utility of (12) depends upon the interval $[b_0, b_1]$ contained within the domain of significant support of $f_{\hat{a}_t}(\hat{a}_t)$ from Eq. (11). As for the knowability issue previously raised regarding A, this will be discussed further into the section.

Now to apply a foundational principle from information theory based on its original application to messages: Consider a statistically independent symbol from a set, for example 'j' ∈ {'s', 'F', 'h', 'j', 'r', '$', 'e', 'P'}. Shannon showed the self-information, in bits, of such an independent symbol issuance in a message confined to the characters in the foregoing set as follows. With the probability of symbol 'j' issuance by a source being Pr('j'), and absolutely independent of any other symbol issuance, its self-information would be absolutely constant, with any issuance yielding self information I_j:

$$I_{\text{j}} \; = \; \log_2 \frac{1}{\mathbf{Pr}(\text{'j'})} \; \text{ bits} \quad (13)$$

Now consider the self information of a data point (cluster member) from untruncated set Q_u having Gaussian PDF designated I_u, and the self information of a data point from truncated set $Q_t \subset Q_u$ designated I_t. The marginal self information for a single data point from a truncated set Q_t and termed ΔI_t will be established as

$$\Delta I_{\text{t}} \; = \; I_{\text{t}} - I_{\text{u}} \quad (14.1)$$

So given the expression (13) of self-information, the marginal self-information for a single data point from a truncated set Q_t having arbitrary PDF would similarly employ the log function[8], and a definition can be proposed based on an analogous inverse probability

$$\Delta I_{\text{t}} \; \stackrel{\text{def}}{=} \; \log_2 \frac{1}{\Pr\left(\hat{A}_{\text{u}} \in Q_{\text{t}} \;\middle|\; |Q_{\text{u}}| \to \infty\right)} \quad (15.1)$$

where the 't' subscript of the foregoing indicating any data point $\hat{A}_t$, and would be applicable for arbitrary $f_{\hat{a}_u}(\hat{a}_u)$, but for now will be applied to rectangular $f_{\hat{a}_u}(\hat{a}_u)$. An obvious requirement is for A

8 The convention here is to apply $\log_{b=2}$ because of the binary magnitude criterion of Eq. (19). Self-information can be specified for any log base with self-information units of *bits* for b = 2 and *nats* for b = *e* .

bracketed by $\{b_0, b_1\}$ -- and for larger intervals and rectangular $f_{\hat{a}_u}(\hat{a}_u)$ and in cases of $B_u = 0$, centered in the interval at the mean of the two bounds for such rectangular PDF. This last feature would be the result of competent, good faith choice of bounds for $\hat{A}_t$ by party X.

The proposal of Def. (15.1) has a basis similar to that proposed by Woodward and Davies in the Summary of an earlier version of Reference (2) antecendently published[9]. To quote the first sentence of the Summary: “The theoretical accuracy and certainty with which range may be determined by radar is obtained quantitatively by applying the principle of inverse probability.” Note that the log arguments of (13) and (15.1) are confined to unity or greater as will be the case for all applications of the log function to come. Now with $f_{\hat{a}_u}(\hat{a}_u)$ modeled as rectangular or quasi-rectangular, (14.1) can be applied to marginal self information designated ΔI_{t_rect} in this rectangular case:

$$\Delta I_{t\_rect} \overset{\text{def}}{=} I_{t\_rect} - I_{u\_rect} \quad (14.2)$$

A special case of interest which will be the defining condition heretofore is when $I_{u\_rect} = 0$, and the parameters of $\hat{A}_u$ and their values with which to assure this condition will be identified further into the development. With this, (14.2) and (15.1) lead to:

$$I_{t\_rect} \overset{\text{def}}{=} \log_2\left[\frac{1}{\Pr\left(\hat{A}_u \in Q_t \;\middle|\; |Q_u| \to \infty\right)}\right]_{I_{u\_rect} = 0} \quad (16)$$

To simplify the subsequent steps, a null value for I_{u_rect} will be assumed, here forward. A probabilistic analog between Eq. (13) and Def. (15.2) can be indicated as follows and so as a special case, consider PDF f_ε unknown except, in keeping with (15.2) quasi-rectangular (but with additional possibly unknown parameters). And importantly, bounds $\{b_0 , b_1\}$ are required in the rectangular PDF case to be within the span of significant support of $f_{\hat{a}_u}(\hat{a}_u)$. And so with prior information as rectangular modeled f_ε, and the conditions of expression (7), the denominator of (15.2) can be modelled:

$$\Pr\left(\hat{A}_u \in Q_t\right) = \Pr\left(b_0 < |\hat{A}_u| < b_1\right) \approx k_0\left[b_1 - b_0\right]_{(b_1 - b_0) \le 1/k_0} \quad (17)$$

9 Woodward, P. M. and Davies, I. L., “A Theory of Radar Information” Phil Mag 41, 1001, 1950

Now with Eq. (17) dimensionless, then k_0, being reciprocally dimensioned with $(b_1 - b_0)$, should contain a factor readily at hand. A liberty will be taken in applying as parameter, $|A|$ as similar to A in Eq. (2). Also the intuitive knowledge that $|A|$ would be, with constant s_ε, monotonic[10] with the I_{rect} from (15.2) and intuited from (10), leads to postulating that k_0 contains a factor $|A|$. The preceding leads to a proposal for a dimensionless truncation constant

$$k_{\text{t}} = \frac{k_0}{|A|} \quad (18)$$

Then Def. (15.2), with equations (17) and (18) incorporated becomes the definition of a new parameter I_{rect}:

$$I_{\text{t\_rect}} = I_{\text{rect}} \overset{\text{def}}{=} \log_2 \frac{|A|}{k_{\text{t}}\left(b_1 - b_0\right)} \quad (15.3)$$

which satisfies monotonicity between $|A|$ and $I_{\text{t\_rect}}$, all other parameters constant.

A somewhat philosophical issue can be introduced and that is whether the numerator of (15.x) should employ $|A|$ as opposed to $|\langle \hat{A}_{\text{u}} \rangle|$. The question is related to the measurand knowability paradox of the problem and relevant to the methodology to be employed, but will not be explored until the end of this section with the discussion of biased estimators.

To determine k_{t} in (15.3), consider the contribution of $\{b_0, b_1\}$ in the Q_{u} cluster truncation for Q_{t} generation, constituting a significant prior information resource in Def. (15.3) and compensating to a degree for information deficit regarding $\boldsymbol{f_\varepsilon}$. So the factor k_{t} in the proposal will be established based on an arbitrary but convenient rule. Suppose it desirable to assign 0 bits self-information to a binary order of magnitude maximum deviation of $\hat{A}_{\text{t}}$ from A, for all $\hat{A}_{\text{t}}$ within this span. In such null case, possible $\hat{A}_{\text{t}}$ values would have a span of

$$\frac{|A|}{2} \le |\hat{A}_{\text{t}}| \le 2|A| \quad (19)$$

with bounds $\{b_0, b_1\}$ so designated as the bounds of (19). And so with the previous is proposed, then

10 Monotonicity and proportionality of these quantities are further referenced below in discussing Eq. (207), putting the requirement into focus.

$$b_0 = \frac{|A|}{2} \quad (20)$$

and $b_1 = 2|A|$ which, both substituted into (15.1), set to zero for null I_{rect} and equated:

$$\log_2 \frac{|A|}{k_t\left(2|A| - \frac{|A|}{2}\right)} = 0 \quad (21)$$

Solving, $k_t = 2/3$ and so Def. (15.2) becomes

$$\Delta I_{\text{rect}} \stackrel{\text{def}}{=} \log_2\left(\frac{|A|}{\frac{2}{3}(b_1 - b_0)}\right) \quad (15.4)$$

Notice the ratio of the bounds in (21), this will be a defining feature useful and referenced further on, and that is for a null valuation of the RHS of (15.4) then:

$$\frac{b_1}{b_0} = 4 \quad (22)$$

Now since (15.3) depends upon $I_{\text{u\_rect}} = 0$, then (15.3) and (22) provide the scenario and a logical argument will be the extent of an explanation: the range of support for rectangular $f_{\hat{a}_u}$ would cover a 4:1 span or greater [from (22)].

As for the Shannon stipulation of surprisal regarding information, Definitions (15.x) apparently fulfill the stipulation as the denominators show. As the (15.3)—(15.4) denominator difference (truncation interval) is widened or narrowed, the result of a measurement $|\hat{A}_u|$ falling within the interval becomes correspondingly less surprising or more so. This suggests the self-information definition as useful in the application at hand.

Now as (15.x) and the resulting (22) have been adopted, so doing places a burden on the reasoning, requiring examining them as conditions of an actual test case. So doing reveals apparent sample (cluster) bias in inequality (19) as seen in cluster Q_t by this truncation, because with a rectangular PDF $|A|$ cannot be placed at the mean between the two bounds in the inequality. This will be noted as an apparent anomaly revealing a but one which is significant for several assumptions offered. This will be examined subsequently but to note and designate this discrepancy it will be stated as a result of possibly a question of good faith, truth or competency (recall the choice of bounds by party X) as in expression (23) below, or possible estimator bias as in expression (24) below, or that the truncation bounds are not established within a usual logical framework. The following are then a mutually exclusive pair designated as the possible origins of 'Discrepancy A':

$$\mathrm{E}\left(|\hat{A}_{\mathrm{t}}| \,\middle\|\, |Q_{\mathrm{t}}| \to \infty\right) \;\approx\; \frac{1}{2}\left(b_0 + b_1\right) \;=\; \frac{1}{2}\left(\frac{|A|}{2} + 2|A|\right) \;\neq\; |A| \quad (23)$$

$$\mathrm{E}\left(\hat{A}_{\mathrm{u}} \,\middle\|\, |Q_{\mathrm{u}}| \to \infty\right) \;\neq\; A \quad (24)$$

Note expression (24) is an outgrowth of the assumption in Eq. (9) with $B_u \neq 0$; and consider it and (23) as alternative causal cases for 'Discrepancy A', as (9) is the feature of an unbiased estimator, and (23) the result of either (1) incompetent or bad faith choice of bounds by party X or (2) party X has knowledge of bias B_u and chooses bounds $\{b_0, b_1\}$ to enhance the self information of individual $\hat{A}_t$ data points. If expression (24) is the causal root of 'Discrepancy A', then the reliance on (19) is logically contaminated by applying the LHS of (9) as substitute for its RHS, with the $B_u = 0$ condition being inoperative. And there will be a proposed consequential link to an imperfection in the heuristic progression when the treatment of estimator bias is undertaken as to its degradation of estimator self-information. So this concern will return as a topic for attention when epistemic logic is introduced, but for now the result of (15.3) will be built upon as a valid conceptual point.

Also of note are three levels of binary attributes of inequality (19): (1) ambiguity-free yes/no inclusion of Q_u data points in Q_t, (2) binary order of magnitude error as the bounds for inclusion, (3) establishment of a pair of parameters $\{b_0, b_1\}$ from perfect knowledge of A in (20) and (22), referred to previously as the 'true value'. This last attribute may seem a triviality but it will be important in subsequent discussion of B_u and the before mentioned accounting for such in self-information determination of $\hat{A}_t$ data points..

Now it would be useful for the progress to replace the rectangular PDF width $(b_1 - b_0)$ in (15.3) with an expression containing the s.d. of a rectangular PDF which will be designated s_{rect}. As the variance of a rectangular PDF of width w is $w^2/12$, then

$$s_{\text{rect}}^2 = \frac{(b_1 - b_0)^2}{12}$$

In this case the s.d. of the truncated cluster would be, from the previous:

$$s_{\text{rect}} = \frac{(b_1 - b_0)}{2\sqrt{3}} \quad (25)$$

Substitution of (b_1 – b_0) in (25) into (15.3) and for designating a new constant k_{rect} for rectangular PDF in the second line following:

$$\begin{aligned} \Delta I_{\text{rect}} &\overset{\text{def}}{=} \log_2\left(\frac{|A|}{\frac{4\sqrt{3}}{3}\left(s_{\text{rect}}\right)}\right) \\ &= \log_2\left(\frac{|A|}{k_{\text{rect}}\, s_{\text{rect}}}\right) \end{aligned} \quad (15.4)$$

The validity of the Def. (15.x) proposal is not assumed at this point but logical inference should be a possible way forward. For example, the yes/no criterion of (19) will be seen in section XIII to resemble a similar criterion to be employed in that section for deriving self-information of a quantity in an important binary decision step with indicator function [Eq. (152)] where the stochastic quantity introduces a quantized timing error, referred to an expected time span. And so the findings in later sections will be employed here with some logical effort.

And as it happens, measurement error can and usually does come from numerous mechanisms, many of which are Gaussian or quasi-Gaussian, and the sum of which will approximate the Gaussian PDF in agreement with the Central Limit Theorem. And so the heuristic Def. (15.x) based on rectangular PDF should be extended to one that accounts for the commonly encountered Gaussian error and this will be a step towards application of the progress to an arbitrary PDF.

And as it turns out, the time-of-arrival (*TOA*) statistics of energy pulse reception (a category to be introduced in later sections) provide for an approach to this endeavor. And a primary reason is (postulated here) that with a typical measurement of signal transit time, the special case of zero transit time (in theory) allows the 'measuring stick' so to say, to be of zero length. And thus the measurement error to be due to a single injected Gaussian variable representing noise of a theoretically known characteristic and thus unbiased as was presumed in the initial assumptions. Notice that with the measurand *A* constrained to be nonzero, this has not been possible so far. And so with *TOA* statistics, a single random variable can model a generalized measurement, itself with a composite of numerous error sources. And so it is possibly the case that a self-information rigorous derivation in the general estimator scenario is not feasible. And this is speculatively because of the inability of the general case to conform to the null expected estimator [E(*TOA*) = 0] scenario applied for example in the final section of the report, and proposed in this paragraph as very useful to the immediate problem of application to generalized estimators, and as such forcing an unbiased (B_u = 0) case as foundational to the progress, and so this will be assumed until specified otherwise.

At this point the introduction of a development from this thesis, Eq. (223), from section IX will provide another model for self-information, termed I_{x_c} of a scalar measure, that of time-of-arrival (*TOA*) of captured signal x_c:

$$I_{xc} = \log_2\left[\frac{\tau_s}{\sqrt{2}\sigma_{t_1}}\right] \quad (223)$$

In Eq. (223) σ_{t_1} is the arrival time resolution, defined as as arrival time t_1 measurement s.d. of signal x_c. The salient point is that the *TOA* is a point in time placed within an intrinsic noise gate interval which is identically the signal duration τ_s (see footnote11, page 14). And with the special case of null transit time, the numerator of (223) exists as an analog of the numerator of (15.4) even though it is not a measurand as is the latter.

Thus is analogous similarity between Def. (15.x) and Eq. (223) and so (223) will be the Gaussian model equivalent to Def. (15.4) with a known s.d. substitution, and with the interval τ_s the substitute for $|A_0|$ (note the intrinsic positive polarity of each). The rationale for this is that Eq. (223) is derived from a process with additive Gaussian noise, and the PDF of the *TOA* detection is itself Gaussian. So as a result, the analogy of (223) to Def. (15.2), itself which would assume a non-Gaussian PDF for ϵ, is, now assuming a Gaussian rather than a rectangular one, and a Gaussian distributed unbiased estimator $\hat{Z}$ for a new nonzero scalar, *Z*, is now introduced with self-information of

$$
\begin{aligned}
I_{\hat{Z}\_\mathrm{Gauss}} &= \log_2\left(\frac{|Z|}{\sqrt{2}\, s_{\hat{Z}}}\right) \\
&= \log_2\left(\frac{|Z|}{k_{\mathrm{gauss}}\, s_{\hat{Z}}}\right)
\end{aligned} \tag{26}
$$

and with the constant in the denominator $k_{\mathrm{gauss}} = \sqrt{2}$ as in Eq. (223).

Element substitution between Def. (15.2) and Eq. (223) as a heuristic for arriving at Eq. (26) for the Gaussian case are summarized in Table 1, a somewhat unconventional alignment map of the factors among the RHS expressions indicated in column headings. It is presented here to aid logical associations moving left to right to generalization in the last column. There is a column for each step of progress and rows for comparison of analogous RHS quantities of each self-information expression. Because of the significant trial-and-error approach employed, Table 1 is an organizational aid for the progress, the last column being the product of the generalization effort.

Table 1: Set Mapping analogous quantities

Expression (PDF model) / Parameter	1 Def. (15.4) (rectangular)	2 Eq. (223) (Gaussian)	3 Eq. (26) (Gaussian)	4 Eq. (32) (arbitrary)
Standard deviation (s.d.)	s_{rect}	σ_{TOA}	s_{gauss}	$s_{\hat{Z}}$
Denominator factor k_{d}	$k_{\mathrm{rect}} = \frac{4\sqrt{3}}{3}$	$k_{\mathrm{gauss}} = \sqrt{2}$	$k_{\mathrm{gauss}} = \sqrt{2}$	$k_{\mathrm{arb}} = \frac{3\sqrt{2}}{\kappa_{\hat{Z}}}$
Dimensioned numerator variable	$\lvert A\rvert$	τ_{s} [11]	$\lvert Z\rvert$	$\lvert Z\rvert$
Skew, γ_1	0	0	0	arbitrary[12]
Kurtosis, κ_{xx}	$\kappa_{\mathrm{rect}} = 9/5$	$\kappa_{\mathrm{gauss}} = 3$	$\kappa_{\mathrm{gauss}} = 3$	$\kappa_{\hat{Z}}$

11 In Eq. (223), τ_{s}, a time interval, is not a measurand magnitude as are the other row factors because of the particular case of arrival time measurement, in which the measurand is a point in time placed within τ_{s}. It is placed in this row because like the scalar measurands, its log magnitude is proportional to self-information determination in the related expression. It is similar to a measurand 'true (absolute) value' in that it is treated as a knowable positive parameter, also it is presumed proportional to a related noise s.d. as it is proportional to noise energy because of its intrinsic noise gate function as discussed immediately following Eq. (68).

12 Skew is at this stage having unknown and likely no relevance to the development.

An elemental view of Table 1 is with column and row headings as set definitions. With this view, then each table entry is an expression factor (possibly intrinsic) representing the intersection of two sets minimum, from its column(s) and its row. For example from the table

$$\{\kappa_{\text{gauss}}\} \;=\; \text{Kurtosis} \cap \text{Eq.}(223) \cap \text{Eq.}(26)$$

with κ_{gauss} not explicitly shown in Eq. (223) or Eq. (26) but will be considered intrinsic to the expressions and this assumption will be useful in a derivation of column 4 - resulting in Eq. (32).

And so the similarity among the parameters of Def. (15.3) and Eq. (26) can be seen to be driven by the more rigorous analog in Eq. (223). The PDF model parameters are included and will be useful for determining the quantities in the last column for an expression to be determined for arbitrary PDF models. Equation (32) in the last column is a general formulation rendered generalized by the inclusion of kurtosis, and is to be derived in the following steps.

And so the similarity among the parameters of Def. (15.3) and Eq. (26) can be seen to be driven by the more rigorous analog in Eq. (223). The PDF model parameters are included and will be useful for determining the quantities in the last column for an expression to be determined for arbitrary PDF models. Equation (32) in the last column is a general formulation rendered generalized by the inclusion of kurtosis, and is to be derived in the following steps.

Since the PDF of A is unknown in Def. (15.2) [except for the general shape], the relationship between $\{b_0, b_1\}$ and s_ε cannot be established; $\{b_0, b_1\}$ is based on a choice by the investigator in possession of prior knowledge. But it will be assumed that s_ε and s_{gauss} are monotonically related to $|A|$, as virtually all real measures are indirect, plus confidence intervals are most frequently dependent monotonically with magnitude of a related measurand.

The interval τ_s, can serve as an intrinsic noise gate in a multiple pass matched filter application, and so as a result, σ_{TOA} is proportional to τ_s, and this is shown by another equation from section X, where :

$$\sigma_{\text{TOA}} \;=\; \frac{\sigma_Z}{m_1} \;=\; \frac{\sigma_Z \tau_s}{\zeta N_{\mathbb{R}}} \quad (207)$$

which establishes monotonicity of σ_{TOA} with the interval τ_s one of the criteria of this section. And so in equations (223) and (26) the numerators are analogous based on the monotonic relationships with measurement s.d. In addition, a final property identification that would further the validation of (26) and that is that its s.d. and that of (223) can take be seen as dependent variables dependent upon the numerators.

It would be natural to speculate as to why the denominator factors in Table 1 are significantly different for the expressions Def. (15.4) and Eq. (223). A hint of a basis would seem to be the parameters in the last two rows and as both expressions exhibit identical skew, the kurtosis measure would seem to offer a basis. And so the comparisons of the two following LHS ratios of parameters from the table can be examined by their product for comparison:

$$\begin{aligned}\left(\frac{k_{\text{gauss}}}{k_{\text{rect}}}\right)\left(\frac{\kappa_{\text{gauss}}}{\kappa_{\text{rect}}}\right) &= \frac{\sqrt{2}}{4\sqrt{3}/3}\,\frac{3}{9/5} \\ &= \frac{5\sqrt{2}}{4\sqrt{3}} = 1.020621\end{aligned} \qquad (27)$$

So with equation (27) there is way forward towards a self-information solution for estimators fitting arbitrary PDF's by the incorporation of kurtosis into Eq. (26). So to do this, the LHS ratios of (27) first line will be set approximately equal after inversion of the rightmost one:

$$\frac{k_{\text{gauss}}}{k_{\text{rect}}} \approx \frac{\kappa_{\text{rect}}}{\kappa_{\text{gauss}}} \quad (28)$$

The suggestion then might be that kurtosis is a major factor in formulating a general form (i.e. applied to arbitrary PDF) for the estimator self-information. So a precept to be proposed is that expression (28), is important information allowing extrapolation of the self-information equations for rectangular and Gaussian distributions to another expression for arbitrary distributions - in similar fashion as a straight line determination by a pair of points.

As close in value as equation (27) is to unity, one could argue that its two LHS ratios are not identical and so signify nothing that can be proven. However they are constructed of integers in one and integers and their square roots in the other. It might also be said that the imperfectly equated quantities of (28) are equated, one can postulate the inequality (23) (Discrepancy 'A') as the source imperfections in the progress and as such preclude the rigorous solution, and this endeavor is one where the widely pursued beauty of mathematics is somewhat obscured. Relatedly, the two distributions studied are of zero skew possibly magnifying the discrepancy.

And to solve the preceding for k_{rect} and in line 2 indicate a severance of its dependence on Def. (15.x) and:

$$\begin{aligned} k_{\text{rect}} &\approx \frac{\kappa_{\text{gauss}} k_{\text{gauss}}}{\kappa_{\text{rect}}} \\ &\approx \frac{3\sqrt{2}}{\kappa_{\text{rect}}} \end{aligned} \quad (29)$$

And because of the speculated heuristic anomaly from the estimator bias of expressions (19) and (21), the RHS of Eq. (29) would be equated (rather than approximated) to a generalized factor contrivance k_{arb}, for denominator factor in column 4 arbitrary PDF model expression. So from Eq. (29) to an arbitrary PDF there is a migration $k_{\text{rect}} \Rightarrow k_{\text{arb}}$ also $\kappa_{\text{rect}} \Rightarrow \kappa_{\hat{Z}}$ for a column 4 generalization in table 1 to arbitrary PDF:

$$\begin{aligned} k_{\text{arb}} &= \frac{\kappa_{\text{gauss}} k_{\text{gauss}}}{\kappa_{\hat{Z}}} \\ &= \frac{3\sqrt{2}}{\kappa_{\hat{Z}}} \end{aligned} \quad (30)$$

And so based on Eq. (30) a new generalized expression is proposed for estimator $\hat{Z}$ self-information $I_{\hat{Z}}$ with Eq. (26) as the model, but with validity for general PDF for measurand Z:

$$I_{\hat{Z}} = \log_2\left(\frac{|Z|}{k_{\text{arb}} s_{\hat{Z}}}\right) \quad (31)$$

with second line of Eq. (30) substitution in Eq. (31):

$$I_{\hat{Z}} = \log_2\left(\frac{\kappa_{\hat{Z}}|Z|}{3\sqrt{2}\, s_{\hat{Z}}}\right) \quad (32)$$

Where Z is measurand 'true value' and to reiterate as a definition: Eq. (32) is the self-information of a single data point and is uniformly descriptive of each and every point in isolation (why 'in isolation' will be apparent in the discussion of entropy below). There is not at present a formal proof of Eq. (32) that can validate the steps that are offered here, and formal proof might not be possible. But consider: the introduction of $\kappa_{\hat{Z}}$ and its appearance in Eq. (32) indicates a 9/5 ~ 3 span (κ_{gauss} = 9/5; κ_{rect} = 3) of the log argument between Gaussian and rectangular PDF application. This seems to be a quite reasonable result in support of the methodology, as an increasing kurtosis value would do what is expected, as increasing the self-information of a single estimation.

To this point the self-information measure has been the focus and the quantification of entropy has not been addressed. A relevant question would be regarding self-information of a cluster $Q_{\hat{Z}}$ of replicated estimator points such as in expression (1) which will be repeated for estimating measurand Z of arbitrary PDF with estimator $\hat{Z}$:

$$\omega_Z : \mathbb{R} \rightarrow Q_{\hat{Z}} \subset \mathbb{R} \quad (1.1)$$

For the following discussion, a condition is assumed, and that is that the stochastic components of each member of $Q_{\hat{Z}}$ are uncorrelated with those of any other data point. The quantification of entropy for the members of $Q_{\hat{Z}}$ would necessitate examination of the possibility of redundant information among them. It is not unreasonable to postulate that as measurement points are generated, there is diminishing surprisal value accrued with each as aggregated. So there is the likely non-applicability of Eq. (32) to the aggregate summation of the individual self-information of members of a specific cluster $Q_{\hat{Z}}$ regarding the self-information, $I_{Q\hat{Z}}$, of the cluster; in other terms:

$$I_{Q\hat{Z}} \neq I_{\hat{Z}} \times |Q_{\hat{Z}}|, \quad \{|Q_{\hat{Z}}| > 1\} \quad (33)$$

But to introduce the entropy of each member, $H_{\hat{Z}}$:

$$H_{\hat{Z}} = \frac{I_{Q\hat{Z}}}{|Q_{\hat{Z}}|} \quad (34)$$

And (34) is verifiable with the employment of Eq. (32) and the standard deviation of the members of $Q_{\hat{Z}}$ to show that the self information of set $Q_{\hat{Z}}$ to be the sum of the individual entropy values of the members. And as the mean incorporates summation divided by cardinality $|Q_{\hat{Z}}|$, in this application the summation is of non-correlated stochastic quantities (the stochastic components of the members) which utilizes the Pythagorean sum, that is the square root of the sum of identical variances or $\sqrt{|Q_{\hat{Z}}| s_{\hat{Z}}^2}$. As such then the mean calculation would be the s.d. of the mean of $Q_{\hat{Z}}$ members:

$$\begin{aligned}\langle s_{\hat{Z}} \rangle_{Q\hat{Z}} &= \frac{1}{|Q_{\hat{Z}}|}\sqrt{|Q_{\hat{Z}}| s_{\hat{Z}}^2} \\ &= \frac{s_{\hat{Z}}}{\sqrt{|Q_{\hat{Z}}|}}\end{aligned} \tag{35}$$

Then the self information of $Q_{\hat{Z}}$ is, applying equations (32) and (35):

$$I_{Q\hat{Z}} = \log_2\left(\frac{\kappa_{\hat{Z}}|Z|\sqrt{|Q_{\hat{Z}}|}}{3\sqrt{2}\,s_{\hat{Z}}}\right) \tag{36}$$

And based on the equations (34) and (36), the entropy of each member of $Q_{\hat{Z}}$ is:

$$H_{\hat{Z}} = \frac{1}{|Q_{\hat{Z}}|}\log_2\left(\frac{\kappa_{\hat{Z}}|Z|\sqrt{|Q_{\hat{Z}}|}}{3\sqrt{2}\,s_{\hat{Z}}}\right) \tag{37}$$

And so the issue regarding redundant information among the elements of $Q_{\hat{Z}}$ is demonstrated by equations (36) and (37), as $I_{Q\hat{Z}}$ is not linear with $|Q_{\hat{Z}}|$.

An interesting conjecture regarding Eq. (36) is the possible accordance between it and the Central Limit Theorem (CLT). The indicator in (36) due to the PDF shape of the underlying population at the basis of the $Q_{\hat{Z}}$ members is $\kappa_{\hat{Z}}$. As cardinality $|Q_{\hat{Z}}|$ is allowed to grow very large, by inspection $\kappa_{\hat{Z}}$ will contribute a vanishing relative contribution to the evaluation of (36) because of the log function. For example if $\kappa_{\hat{Z}} = 9/5$ (as in the rectangular PDF) in (36) then $I_{Q\hat{Z}}$ evaluation would be essentially equivalent for $\kappa_{\hat{Z}} = 3$ in the Gaussian case as $|Q_{\hat{Z}}|$ grows very large. A closer agreement between the two cases as $|Q_{\hat{Z}}|$ grows large, could be seen as compatibility between Eq. (36) and the CLT. As the quantity within parentheses of Eq. (36) and (37) becomes large with increasing measurement count

$|Q_{\hat{Z}}|$, alternating between $\kappa_{\hat{Z}} = 3$ and $\kappa_{\hat{Z}} = 9/5$ generates a smaller and smaller marginal difference in the evaluation of the expression.

Now it would seem effort could be directed towards biased estimators with an attempt to account for bias in the self information of scalar estimators. As a preliminary discussion, it may help to explore the topic of epistemic logic further because the aforementioned Eq (23) "Discrepancy 'A' will return as a principle to aid examination of biased estimators and how to treat them as a continuation of the topic. A strict adherence to expression (15.3) would require specific knowledge of *A,* one may even insist on unrestricted, or perfect knowledge of *A*. The knowability thesis of epistemic logic will be the governing principle of such perfect knowledge.

Since measurement literature refers to the "true value" quantity, here it seems valid to explore the relationship of the "true value" to a truth knower, or the knower of *A* in expression (15.4) and to any party with no knowledge of *A,* here to be personified as party Y. According to the knowability thesis (i.e. all truths are in theory knowable), it should be reasonable to postulate a knower of *A* and some statistical properties of estimator $\hat{A}$, and for this designate the aforementioned party X. Party X can participate in an investigation with measurements of *A* and can obtain the measurements $\hat{A}_u$ in Q_u.and so from the data in Q_u recognizes the rectangularity of the related PDF. The bias B_u of estimator $\hat{A}$ is known to X and not to Y and so knowledge of *A* and B_u constitute prior information accessible to X only. There is a degradation of $\hat{A}$ self information with increasing estimator bias magnitude, up to a bias extremity B_{ext} which is a value at which bias magnitude confers zero estimator self information, which is known to X and is to be determined by this exercise. Clearly X can arithmetically approximate $\hat{A}$ bias from the knowledge of *A* and its difference from $\langle \hat{A}_u \rangle$ but is granted perfect knowledge of bias B_u nevertheless.

Suppose that parties X and Y are mutually committed to truth. So X could determine bounds $\{b_0, b_1\}$ for truncation and generate cluster Q_t and pass the data in Q_t to party Y, or conversely choose not to truncate the data. Party Y knows that X is a knower of *A* and bias B_u and that with the commitment to truth, that X would not choose bounds as to intentionally mislead Y, which is to say degrade the self-information of $\hat{A}_t$. However X is under no obligation to supply the value of *A*, either explicitly or as to be approximated by $\langle \hat{A}_t \rangle$, only under obligation to supply truncated values in Q_t with self-information value not less than untruncated Q_u. This last requirement may seem contradictory at first, but can be